

# Limit order books

Martin D. Gould,<sup>1,2,3,\*</sup> Mason A. Porter,<sup>1,3</sup> Stacy Williams,<sup>4</sup> Mark McDonald,<sup>4</sup> Daniel J. Fenn,<sup>4</sup> and Sam D. Howison<sup>1,2</sup>

<sup>1</sup>*Oxford Centre for Industrial and Applied Mathematics,  
Mathematical Institute,  
University of Oxford,  
Oxford OX1 3LB,  
UK*

<sup>2</sup>*Oxford-Man Institute of Quantitative Finance,  
University of Oxford,  
Oxford, OX2 6ED,  
UK*

<sup>3</sup>*CABDyN Complexity Centre,  
University of Oxford,  
Oxford OX1 1HP,  
UK*

<sup>4</sup>*FX Quantitative Research,  
HSBC Bank, London E14 5HQ,  
UK*

Limit order books are used to match buyers and sellers in more than half of the world's financial markets, and have been studied extensively in several disciplines during the past decade. This survey highlights the many insights from the wealth of empirical and theoretical studies that have been conducted, and the numerous unsolved problems that remain. We illustrate the differences between observations from empirical studies of limit order books and the models that attempt to replicate them. In particular, many modelling assumptions are poorly supported by data and several well-established empirical facts have yet to be reproduced satisfactorily by models. By examining existing models of limit order books, we identify some key unresolved questions and difficulties currently facing researchers of limit order trading.

## CONTENTS

I. Introduction	2	3. The Hurst exponent	15
II. A Mathematical Description of Limit Order Trading	2	4. Estimators of $H$	15
A. Preliminaries	3	IV. Empirical Observations in Limit Order Markets	16
B. Price changes in limit order books	5	A. Order size	16
C. Orders: the building blocks of a limit order book	6	B. Relative price	17
D. The rise in popularity of limit order books	6	C. Limit order cancellations	18
III. Limit Order Book Challenges	7	D. Mean relative depth profile	18
A. Different perspectives	7	E. Volatility	18
B. State-space complexity	7	F. Conditional event frequencies	19
C. Feedback	8	1. Order size	19
D. Coupling between $b(t)$ and $a(t)$	8	2. Relative price	19
E. Quantifying patience	8	3. Arrival rates	20
F. Priority	8	4. Cancellations	21
G. Incomplete sampling and hidden liquidity	9	5. Price movements	21
H. Volatility	10	6. Incoming order prices	21
1. Model-based estimates of volatility	11	7. Order flow	22
2. Model-free estimates of volatility	11	8. Limit order book state	22
3. Time window effects	11	G. Market impact and price impact	22
I. Resolution parameters	12	1. Instantaneous price impact	24
J. Bilateral trade agreements	12	2. Permanent price impact	25
K. Opening and closing auctions	13	3. Market impact	25
L. Power laws	13	H. Stylized facts	26
M. Long-Range Correlations	14	1. Heavy-tailed return distribution	26
1. Long- and short-memory processes	14	2. Autocorrelation of returns	27
2. Practical considerations	15	3. Long memory	27
		V. Limit Order Book Models	28
		A. Perfect-rationality approaches	28
		1. Cut-off strategies	28
		2. Fundamental values and informed traders	29
		3. Minimizing market impact	30
		B. Zero-intelligence approaches	31
		1. Model framework	31

\* Corresponding author: gouldm@maths.ox.ac.uk

2. Random-walk diffusion models	31
3. Discrete-time models	32
4. Continuous-time models	32
5. Beyond zero intelligence	34
C. Agent-based models	35
VI. Discussion	36
Acknowledgments	38
References	38
A. Table of Empirical Studies	43

## I. INTRODUCTION

It is an age-old problem to determine the price at which to conduct a trade. In the highly competitive and relentlessly fast-paced markets of today’s financial world, *limit order books* (LOBs) are used to match buyers and sellers in more than half of the world’s financial markets (Roşu, 2009). Euronext, the Australian Securities Exchange, and the Helsinki, Hong Kong, Swiss, Tokyo, Toronto, Vancouver, and Shenzhen Stock Exchanges all now operate as pure limit order markets (Gu *et al.*, 2008b; Luckock, 2001); the New York Stock Exchange (NYSE), NASDAQ, and the London Stock Exchange (LSE) (Cont *et al.*, 2010) all operate a bespoke hybrid limit order system. Thanks to technological advances, traders around the globe now have real-time access to the current LOB, providing buyers and sellers alike “the ultimate microscopic level of description” (Bouchaud *et al.*, 2002).

There are obvious practical advantages to understanding the dynamics of the LOB. These include gaining clearer insight into the relative financial merits of different kinds of orders in given situations (Harris and Hasbrouck, 1996); optimal order execution strategies (Obizhaeva and Wang, 2005); and market impact minimization (Eisler *et al.*, 2010). By providing researchers with more detailed information about order flow than has ever been available before, the study of LOBs has helped illuminate possible causes for well-established empirical regularities that have been observed across a wide range of markets (Bouchaud *et al.*, 2009; Farmer and Lillo, 2004). Furthermore, a LOB is a prime example of a *complex system* (Mitchell, 2009), where global phenomena arise from the behaviour of many interacting agents when the system throughput becomes sufficiently large. The unusually rich and high-quality historic data that is available from some LOBs make them a suitable testing ground for a wide variety of ideas from the complex systems literature, such as universality, scaling, and emergence. In this survey, we discuss some of the key ideas that have emerged from the analysis and modelling of limit order trading in recent years. We also examine existing models of limit order trading and discuss their strengths and limitations.

This survey is preceded by a number of other survey articles that each focus on particular aspects of the mechanism. Friedman (2005) reviewed early studies of “double auction” style trading (of which the LOB is an example). Parlour and Seppi (2008) addressed the economic and theoretical aspects of the trading mechanism. Bouchaud *et al.* (2009) assessed the current understanding of price formation in limit order markets. Chakraborti *et al.* (2011a,b) examined the role of econophysics in understanding LOB behaviour. In light of these studies, we do not focus heavily on these aspects here.

Analyses of LOBs have taken a variety of starting points, drawing on ideas from economics, physics, mathematics, statistics, and psychology. Unsurprisingly, there is no clear consensus on the best approach. This point is exemplified by the contrast between the “bottom-up” approach normally taken in the economics literature, in which models focus on the behaviour of individual market participants and present the LOB as a sequential game (Foucault, 1999; Parlour, 1998; Roşu, 2009), with the approach in the physics literature, in which order flows are treated as random and techniques from statistical mechanics are employed to explore LOB equilibria (Challet and Stinchcombe, 2001; Cont *et al.*, 2010; Smith *et al.*, 2003). In the current paper, we discuss developments in both the economics and physics literatures, while emphasizing the aspects of limit order trading most relevant to practitioners.

The remainder of the survey is organized as follows. In Section II, we discuss several formal definitions related to limit order trading, in order to formulate a mathematical description of the process. In Section III, we discuss a variety of practical aspects of limit order trading and examine the mathematical difficulties that arise from attempting to quantify them. In Section IV, we examine the important role of empirical studies in deepening understanding of the LOB, highlighting both consensus and disagreement within the literature. We examine a selection of models in Section V. Finally, Section VI contains our conclusions and discussions of what we believe to be the key unresolved issues in the field.

## II. A MATHEMATICAL DESCRIPTION OF LIMIT ORDER TRADING

In this section, we formulate a precise description of limit order trading that is intended to describe the key aspects common to most limit order markets. Of course, individual exchanges and trading platforms each operate under slight variations of these core principles. Harris (2003) provides a comprehensive review of specific details governing particular exchanges, so we do not focus heavily on them here.

## A. Preliminaries

Before limit order trading grew in popularity, most financial trades took place in *quote-driven* marketplaces, where a handful of large *market makers* centralize “buy” and “sell” orders by publishing the prices at which they are willing to buy and sell the asset being traded. The sell price will always be higher than the buy price, allowing the market maker to earn a profit in exchange for providing *liquidity* to the market<sup>1</sup> and taking on the risks of market making. Such risks include acquiring an undesirable inventory position that later has to be unwound and being exposed to *adverse selection* (i.e., encountering other market participants who have better information about the value of the asset and who can therefore make a profit by buying or selling, often repeatedly, with the market maker (Parlour and Seppi, 2008)). Any other market participants who want to buy or sell the asset only have access to the prices made publicly available by these market makers, and the only action available to a market participant who wishes to buy (respectively, sell) the asset is to immediately buy (respectively, sell) at one of the prices that the market makers have made available. The purchasing and sale of tickets via ticket touts is an example of a quote-driven market in action.

A *limit order market* is much more flexible because every market participant has the option of posting buy (respectively, sell) *orders*:

**Definition.** An order  $x = (p_x, \omega_x)$  at price  $p_x$  and of size  $\omega_x > 0$  (respectively,  $\omega_x < 0$ ) is a commitment to sell (respectively, buy) up to  $|\omega_x|$  units of the asset being traded at a price no less than (respectively, no greater than)  $p_x$ .

The units of order size are set by the following:

**Definition.** The lot size, denoted  $\sigma$ , is the smallest amount of the asset that can be traded in the market. Furthermore, all orders<sup>2</sup> must arrive with a size  $\omega_x \in \{k\sigma \mid k \in \mathbb{Z}\}$ .

The units of price are set by the following:

**Definition.** The tick size  $\delta p$  is the smallest price interval between different orders that is permissible in the market. Furthermore, all orders must arrive with a price that is specified to the accuracy of  $\delta p$ .

<sup>1</sup> Liquidity is difficult to define formally. Kyle (1985) instead identified the three key properties of a liquid market to be tightness (“the cost of turning around a position over a short period of time”), depth (“the size of an order-flow innovation required to change prices a given amount”) and resiliency (“the speed with which prices recover from a random, uninformative shock”).

<sup>2</sup> In some real markets there are two lot-size parameters: a minimum size  $\sigma$  and an increment  $\varepsilon$ . In such markets, sell orders must arrive with a size  $\omega_x \in \{\sigma + k\varepsilon \mid k = 0, 1, 2, \dots\}$  and buy orders must arrive with a size  $\omega_x \in \{-(\sigma + k\varepsilon) \mid k = 0, 1, 2, \dots\}$ . For simplicity, we assume  $\sigma = \varepsilon$ .

For example, if  $\delta p = \$0.00001$ , then the largest permissible order price that is strictly less than \$1.00 is \$0.99999, and any order would have to be submitted at a price with exactly 5 decimal places.

**Definition.** The lot size  $\sigma$  and the tick size  $\delta p$  are collectively called the *LOB’s* resolution parameters.

When a buy (respectively, sell) order is submitted, a *LOB’s trade-matching algorithm* checks whether it is possible to perform a matching to some other previously submitted sell (respectively, buy) order. If so, the matching occurs immediately. If not, the newly submitted order becomes *active*; it remains active until either it becomes matched to another incoming sell (respectively, buy) order or it is cancelled.

**Definition.** An active order at time  $t$  is an order that has been submitted at some time  $t' \leq t$ , but has not been matched or cancelled by time  $t$ .

Cancellation usually occurs because the owner of an order no longer wishes to offer a trade at the stated price, but rules governing the marketplace can also lead to the cancellation of orders if they remain unmatched for a specified period of time, or at certain times of day. For example, on the electronic limit order trading platform Hotspot FX, all active orders are cancelled at 5pm EST to prevent too large an accumulation of active orders over time (Gould *et al.*, 2011).

It is precisely the active orders in a market that make up the *LOB*:

**Definition.** The *LOB*  $L(t)$  is the set of all active orders in the market at time  $t$ .

The *LOB* is then a set of queues, with each queue consisting of active buy or sell orders at the specified price. The concepts of the *bid price*, *ask price*, *mid price*, and *bid-ask spread*, common to much of the finance literature, can be made specific in the context of limit order trading:

**Definition.** The bid price at time  $t$ , denoted  $b(t)$ , is the highest stated price among active buy orders in the *LOB*  $L(t)$ ,

$$b(t) = \max_{\{x \in L(t) \mid \omega_x < 0\}} p_x.$$

**Definition.** The ask price at time  $t$ , denoted  $a(t)$ , is the lowest stated price among active sell orders in the *LOB*  $L(t)$ ,

$$a(t) = \min_{\{x \in L(t) \mid \omega_x > 0\}} p_x.$$

**Definition.** The bid-ask spread at time  $t$ , denoted  $s(t)$ , is the difference between the ask price and the bid price at time  $t$ :  $s(t) = a(t) - b(t)$ .

**Definition.** The mid price at time  $t$ , denoted  $m(t)$ , is the average of the ask price and the bid price at time  $t$ :  $m(t) = \frac{a(t)+b(t)}{2}$ .

In a limit order market,  $b(t)$  is the highest price at which it is possible to sell immediately at least the lot size of the asset being traded, and  $a(t)$  is the lowest price at which it is possible to buy immediately at least the lot size of the asset being traded, at time  $t$ .

Sometimes it is more helpful to consider prices relative to  $b(t)$  and  $a(t)$ , rather than actual prices:<sup>3</sup>

**Definition.** For a given price  $p$ , the bid-relative price  $\Delta_b(p)$  is the difference between the bid price and the given price:  $\Delta_b(p) = b(t) - p$ .

**Definition.** For a given price  $p$ , the ask-relative price  $\Delta_a(p)$  is the difference between the given price and the ask price:  $\Delta_a(p) = p - a(t)$ .

Notice the difference in signs between the two above definitions:  $\Delta_b(p)$  measures the distance that  $p$  is “behind”  $b(t)$ , i.e., how much smaller  $p$  is than  $b(t)$ ;  $\Delta_a(p)$  measures the distance that  $p$  is “behind”  $a(t)$ , i.e., how much larger  $p$  is than  $a(t)$ .

Often it is desirable to compare orders on the bid side and the ask side of the limit order book. In these cases, the concept of a single *relative price of an order* is useful:

**Definition.** For a given order  $x = (p_x, \omega_x)$ , the relative price of the order, denoted  $\Delta_x$ , is defined as

$$\Delta_x = \begin{cases} \Delta_b(p_x), & \text{if the order is a buy order,} \\ \Delta_a(p_x), & \text{if the order is a sell order.} \end{cases}$$

Arriving orders may then be partitioned according to their relative price:

- Any order  $x$  that arrives with a relative price  $\Delta_x \geq 0$  will not cause an immediate matching, instead becoming an active order “inside the book”.
- Any order  $x$  that arrives with a relative price  $-s(t) < \Delta_x < 0$  will also not cause an immediate matching, instead becoming an active order “inside the spread”.
- Any order  $x$  that arrives with a relative price  $\Delta_x \leq -s(t)$  will cause an immediate matching.

In order to assess the state of  $L(t)$ , most market participants view the *depth profile* or *relative depth profile* of the LOB:

<sup>3</sup> In the existing literature, a wide range of different naming and sign conventions are used by different authors to describe slightly different definitions of the concept of relative price. Here, we introduce the explicit distinction between bid-relative price and ask-relative price in an attempt to remove any potential confusion.

**Definition.** The bid-side depth available at price  $p$  and at time  $t$ , denoted  $n_b(p, t)$ , is the total size of all active buy orders at price  $p$ ,

$$n_b(p, t) = \sum_{\{x \in L(t) | p_x = p, \omega_x < 0\}} \omega_x.$$

The ask-side depth available at price  $p$  and at time  $t$ , denoted  $n_a(p, t)$ , is the total size of all active sell orders at price  $p$ ,

$$n_a(p, t) = \sum_{\{x \in L(t) | p_x = p, \omega_x > 0\}} \omega_x.$$

The depth available is often stated in multiples of the lot size. Notice that because  $\omega_x < 0$  for buy orders and  $\omega_x > 0$  for sell orders, it follows that  $n_b(p, t) \leq 0$  and  $n_a(p, t) \geq 0$  for all prices  $p$ .

**Definition.** The bid-side depth profile of the LOB at time  $t$  is the set of all ordered pairs  $(p, n_b(p, t))$ . The ask-side relative depth profile of the LOB at time  $t$  is the set of all ordered pairs  $(p, n_a(p, t))$ .

**Definition.** The mean bid-depth available at price  $p$  between times  $t_1$  and  $t_2$ , denoted  $\bar{n}_b(p, t_1, t_2)$ , is

$$\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} n_b(p, t) dt.$$

The mean ask-depth available at price  $p$  between times  $t_1$  and  $t_2$ , denoted  $\bar{n}_a(p, t_1, t_2)$ , is defined similarly, using the ask-side depth available.

Because  $b(t)$  and  $a(t)$  themselves vary over time, it is rarely illuminating to consider the depth available at a specific price over time. However, relative pricing provides a useful alternative:

**Definition.** The bid-side relative depth available at price  $p$  and at time  $t$ , denoted  $N_b(p, t)$ , is the total size of all active buy orders with relative price  $p$ ,

$$N_b(p, t) = \sum_{\{x \in L(t) | \Delta_x = p, \omega_x < 0\}} \omega_x.$$

The ask-side relative depth available at price  $p$  and at time  $t$ , denoted  $N_a(p, t)$ , is the total size of all active sell orders with relative price  $p$ ,

$$N_a(p, t) = \sum_{\{x \in L(t) | \Delta_x = p, \omega_x > 0\}} \omega_x.$$

**Definition.** The bid-side relative depth profile of the LOB at time  $t$  is the set of all ordered pairs  $(\Delta_x, N_b(\Delta_x, t))$ . The ask-side relative depth profile of the LOB at time  $t$  is the set of all ordered pairs  $(\Delta_x, N_a(\Delta_x, t))$ .

**Definition.** The mean bid-depth available at relative price  $p$  between times  $t_1$  and  $t_2$ , denoted  $\bar{N}_b(p, t_1, t_2)$ , is

$$\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} N_b(p, t) dt.$$

The mean ask-depth available at relative price  $p$  between times  $t_1$  and  $t_2$ , denoted  $\bar{N}_a(p, t_1, t_2)$ , is defined similarly, using the ask-side relative depth available.

**Definition.** The mean bid-side relative depth profile between times  $t_1$  and  $t_2$  is the set of all ordered pairs  $(\Delta_x(p), \bar{N}_b(\Delta_x(p), t_1, t_2))$ . The mean ask-side relative depth profile between times  $t_1$  and  $t_2$  is the set of all ordered pairs  $(\Delta_x(p), \bar{N}_a(\Delta_x(p), t_1, t_2))$ .

The relative depth profile provides no information about the actual prices at which trades occur, nor does it contain information about  $s(t)$  or  $m(t)$ . However, order arrival rates have been widely observed to depend on relative prices and not actual prices (see, e.g., (Biais *et al.*, 1995; Bouchaud *et al.*, 2002; Potters and Bouchaud, 2003; Zovko and Farmer, 2002)), so it is common to consider the relative depth profiles and  $b(t)$  and  $a(t)$  together, in order to gain the full picture of LOB dynamics.

It is often also useful to study the properties of all active limit orders together.

**Definition.** The total depth available on the bid side of the LOB at time  $t$ , denoted  $\hat{N}_b(t)$ , is the total size of all active buy orders at time  $t$ ,

$$\hat{N}_b(t) = \sum_{\{x \in L(t) | \omega_x < 0\}} \omega_x.$$

The total depth available on the ask side of the LOB at time  $t$  is defined similarly:

$$\hat{N}_a(t) = \sum_{\{x \in L(t) | \omega_x > 0\}} \omega_x.$$

**Definition.** The total depth available in the limit order book  $L(t)$ , denoted by  $\hat{N}(t)$ , is given by

$$\hat{N}(t) = |\hat{N}_b(t)| + \hat{N}_a(t).$$

Figure 1 shows a schematic of a LOB at some instant in time, illustrating the definitions in this section. The horizontal lines within the blocks at a given price level denote how the depth available at that price is made up from different active orders.

Time series of prices are a common object of study in the empirical literature on LOBs. As discussed in Section IV.H, it is a recurring theme in these studies that the behaviour exhibited by such time series depends heavily on how they are constructed. For example, consider the time series  $m(t_1), \dots, m(t_n)$ , for some set of times  $t_1, \dots, t_n$ :

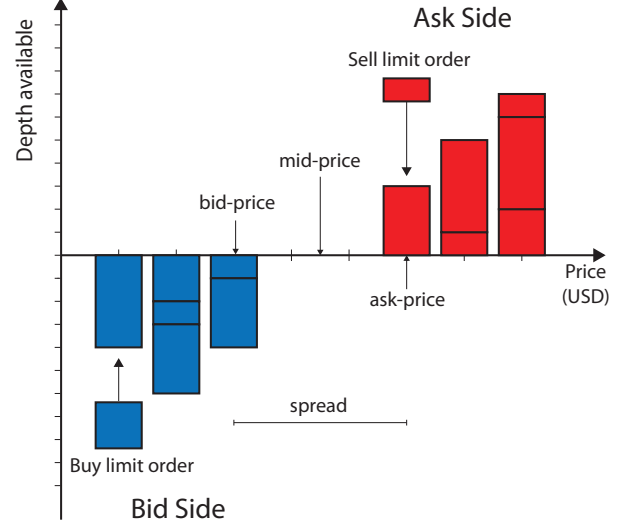


FIG. 1 Schematic of a LOB

- When the  $t_i$  are regularly spaced in time, with an elapsed time of  $\tau$  seconds between them, such a time series is said to be constructed on a  $\tau$ -second timescale.
- When the  $t_i$  are chosen to correspond to the times of arrivals and departures of orders from the LOB, the  $t_i$  are irregularly spaced in time. Such a time series is said to be constructed on an *event-by-event* timescale.
- When the  $t_i$  are chosen to correspond to the times of trades (i.e., matchings in the LOB), the  $t_i$  are also irregularly spaced in time. Such a time series is said to be constructed on a *trade-by-trade* timescale.

## B. Price changes in limit order books

The rules that govern matchings in LOBs dictate how prices evolve through time in these markets. The arrival of an order with a relative price  $\Delta_x \geq 0$  will never cause either of  $b(t)$  or  $a(t)$  to change. When a buy (respectively, sell) order with a relative price  $-s(t) < \Delta_x < 0$  arrives, it will become the active buy order with highest (respectively, active sell order with lowest) price in the LOB, and  $b(t)$  will increase (respectively,  $a(t)$  will decrease) to the price of this order. Whether or not the occurrence of a matching causes  $b(t)$  (respectively,  $a(t)$ ) to change will depend on  $n(a(t), t)$  (respectively,  $n(b(t), t)$ ) and the size of the incoming order that triggered the matching. In particular, the new bid price immediately after the arrival of a sell order  $x = (p_x, \omega_x)$  with a relative price

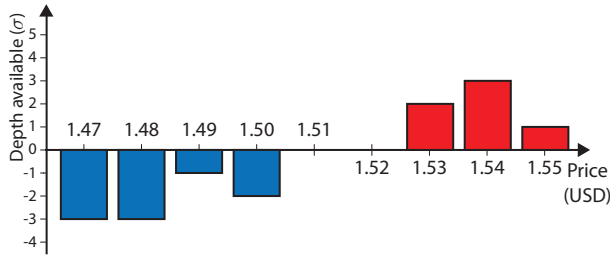


FIG. 2 An example LOB

Event	Values after event (\$)			
	$b(t)$	$a(t)$	$m(t)$	$s(t)$
Initial Values	1.50	1.53	1.515	0.03
Buy order, size 3, price \$1.48	1.50	1.53	1.515	0.03
Buy order, size 3, price \$1.51	1.51	1.53	1.52	0.02
Buy order, size 3, price \$1.55	1.50	1.54	1.52	0.04
Buy order, size 5, price \$1.55	1.50	1.55	1.525	0.05
Sell order, size 4, price \$1.54	1.50	1.53	1.515	0.03
Sell order, size 4, price \$1.52	1.50	1.52	1.51	0.02
Sell order, size 4, price \$1.47	1.48	1.53	1.515	0.05
Sell order, size 4, price \$1.50	1.49	1.50	1.495	0.01

TABLE I How each specified market event would affect prices if it occurred immediately after the initial LOB state displayed in Figure 2

$\Delta_x < -s(t)$  is

$$\max(p_x, q), \text{ where } q = \arg \max_{r'} \sum_{r=r'}^{b(t)} |n(r, t)| > \omega_x. \quad (1)$$

Similarly, the new ask price immediately after the arrival of a buy order  $x$  of size  $\omega_x < 0$  with a relative price  $\Delta_x < -s(t)$  and an actual price  $p_x$  is

$$\min(p_x, q), \text{ where } q = \arg \min_{r'} \sum_{r=a(t)}^{r'} n(r, t) > |\omega_x|. \quad (2)$$

Table I lists a number of possible market events, along with the changes that they would cause to  $b(t)$ ,  $a(t)$ ,  $m(t)$ , and  $s(t)$ , if they were to occur in the LOB displayed in Figure 2.

Price changes in markets are commonly studied via the concept of *returns*:

**Definition.** The bid-price return between times  $t_1$  and  $t_2$ , denoted  $R_b(t_1, t_2)$ , is the change in the bid price between time  $t_1$  and time  $t_2$ :  $R_b = \frac{b(t_2) - b(t_1)}{b(t_1)}$ . The ask-price return between times  $t_1$  and  $t_2$ ,  $R_a(t_1, t_2)$ , and the mid-price return between times  $t_1$  and  $t_2$ ,  $R_m(t_1, t_2)$ , are defined similarly.

**Definition.** The bid-price logarithmic return between times  $t_1$  and  $t_2$ , denoted  $r_b(t_1, t_2)$ , is the natural logarithm of the ratio of the bid price at time  $t_2$  and the bid price at time  $t_1$ :  $r_b = \log \left( \frac{b(t_2)}{b(t_1)} \right)$ . The ask-price logarithmic return between times  $t_1$  and  $t_2$ ,  $r_a(t_1, t_2)$ , and the mid-price logarithmic return between times  $t_1$  and  $t_2$ ,  $r_m(t_1, t_2)$ , are defined similarly.

### C. Orders: the building blocks of a limit order book

Although LOBs provide a far greater level of flexibility to market participants than do quote-driven markets, the actions of market participants in a limit order market can always be expressed solely in terms of the submission or cancellation of orders of the lot size. For example, an impatient market participant who immediately sells  $4\sigma$  units of the asset being traded in the LOB displayed in Figure 2 can be thought of as submitting 2 sell orders of size  $\sigma$  at the price \$1.50, 1 sell order of size  $\sigma$  at the price \$1.49, and 1 sell order of size  $\sigma$  at the price \$1.48. Similarly, a patient market participant who posts a sell order of size  $4\sigma$  at the price \$1.55 can be thought of as submitting 4 sell orders of size  $\sigma$ , each at this price.

In almost all the published literature on LOBs, the following terminology has been adopted. Orders that result in an immediate matching upon submission are known as *market orders*. Orders that do not, instead becoming active orders, are known as *limit orders*.<sup>4</sup> However, it is important to recognize that this terminology is only used to emphasize whether an incoming order triggers an immediate matching or not. There is no fundamental difference between a limit order and a market order.

Some limit order markets allow impatient market participants to specify that they wish to submit a buy (respectively, sell) market order, without explicitly specifying a price. Instead, the market participant specifies only a size, and the matching algorithm sets the price of the order appropriately to initiate the required matching(s).

### D. The rise in popularity of limit order books

Glosten (1994) argued that limit order markets are an effective way for “patient” market participants to provide liquidity to “less patient” market participants, even in situations where liquidity is scarce. He also argued that limit order markets are immune to competition by other exchange set-ups (such as quote-driven marketplaces, as described above). Luckock (2003) concluded that the volume of trade in a limit order market would always exceed

<sup>4</sup> Some practitioners use the terms “aggressive orders” and “resting orders”, respectively, but this terminology is far less common in the published literature.

that of a Walrasian market,<sup>5</sup> given the same underlying supply and demand. Foucault *et al.* (2005) argued that the popularity of limit order markets was due in part to their ability to allow impatient market participants to demand “immediacy”, while simultaneously allowing patient market participants to supply it to those who will later require it.

### III. LIMIT ORDER BOOK CHALLENGES

In this section, we discuss some of the challenges that LOBs present researchers. In particular, we discuss technical issues associated with the study of empirical LOB data, and present several challenges inherent in modelling a LOB within the general framework set out in Section II.

#### A. Different perspectives

In order to construct a useful model of a LOB, certain assumptions must inevitably be made. One such assumption regards the reason that order flows exist at all. In much of the mainstream economics literature, order submission is motivated by the assumption that “perfectly rational” agents attempt to maximize their “utility” by making trades in markets driven by “information” (Parlour and Seppi, 2008). However, this methodology has come under increasing scrutiny. For example, Gode and Sunder (1993) highlighted how utility maximisation is often inconsistent with direct observations of individual behaviour, and Smith *et al.* (2003) noted that the frequency with which existing active orders are modified is far too low to account for the arrival of every new piece of unanticipated information. Therefore, *perfect rationality* must be viewed as an extreme modelling assumption that is made in order to provide a framework within which calculations can be performed.

At the other extreme lies the *zero-intelligence* approach, in which aggregated order flows are assumed to be governed by a specified stochastic process, rather than decomposed into their constituent actions and motivated by individual agents attempting to maximize some personal utility function (Cont *et al.*, 2010; Daniels *et al.*, 2002; Smith *et al.*, 2003). Much like perfect rationality, the zero-intelligence approach is an extreme simplification that is inconsistent with empirical observations. For example, Boehmer *et al.* (2005) found detectable changes in order flows when the NYSE increased the amount of

information about the LOB that was made available to market participants in real time. Furthermore, Bortoli *et al.* (2006) noted that when the Sydney Futures Exchange implemented a similar change, the depth available at the best prices became smaller and market orders for quantities larger than those offered at the best prices became more frequent.<sup>6</sup> This suggests that market participants were using such information when deciding how to interact with the LOB, and were therefore not acting with zero intelligence. However, zero intelligence has the appeal of being an easily quantifiable concept, and leads to falsifiable predictions that are testable without the need for a series of auxiliary assumptions. It is, therefore, a useful starting point for building models.<sup>7</sup>

Between the two extremes of perfect rationality and zero intelligence lies a broad range of other approaches that make weaker assumptions about market participants’ behaviour and order flows, at the cost of resulting in models that are more difficult to study. Many such models rely exclusively on Monte-Carlo simulation to produce output. Although such simulations still motivate interesting observations, it is often difficult to trace exactly how specific model outputs are affected by the input parameters, whereas in analytical work such dependence is explicit throughout.

#### B. State-space complexity

It is a well-established empirical fact that order flows depend on both  $L(t)$  and on recent order flows (Biais *et al.*, 1995; Ellul *et al.*, 2003; Hall and Hautsch, 2006; Hollifield *et al.*, 2004; Lo and Sapp, 2010; Sandås, 2001). From a perfect-rationality perspective, this can be thought of as market participants reacting to the changing state of the market; from a zero-intelligence perspective, this can be thought of as order flow rates depending on  $L(t)$  and on recent order flows. From either perspective, a key task is to uncover the structure of such conditional behaviour, either to understand what information market participants evaluate when making decisions or to quantify the conditional structure of order flows.

However, the state space is huge: if there are  $P$  different choices for price, the state space of the depth profile is  $\mathbb{Z}^P$ . This makes it very difficult to investigate such de-

<sup>5</sup> A Walrasian market is one in which all market participants send their desired buy or sell orders to a specialist, who then determines the market value of the asset by selecting the price that will maximise the volume of trade.

<sup>6</sup> Several other studies have investigated such changes (see, e.g., (Boehmer *et al.*, 2005; Madhavan *et al.*, 2005; Mizrahi, 2008)), but they have all been based on hybrid limit order markets in which (even before the change) some market participants had access to more information about the LOB than others.

<sup>7</sup> We explore in Section V how some authors have attempted to quantify perfect rationality for modelling purposes and discuss the often highly unrealistic assumptions that such formulations require in order to be empirically tested. A more detailed treatment can be found in (Foucault *et al.*, 2005).

pendences, as the number of variables upon which to condition is so large. Therefore, a key task in LOB modelling is to find a way to simplify the evolving, high-dimensional state space, while retaining the important features. Although some suggestions for such dimensionality reduction have been made (see, e.g., (Cont and de Larrard, 2011; Eliezer and Kogan, 1998; Smith *et al.*, 2003)), there is no consensus about a simplified state space upon which a very general LOB model can be constructed.

### C. Feedback

In addition to market participants' actions depending on  $L(t)$ , it is also clearly true that the state of the LOB  $L(t)$  is entirely dependent on market participants' actions. These two mutual dependences form a feedback loop between  $L(t)$  and market participant behaviour that makes LOB modelling very difficult.

### D. Coupling between $b(t)$ and $a(t)$

As described in Section II.C,  $b(t)$  determines the boundary condition for sell limit order placement because any sell order placed at or below  $b(t)$  will at least partially match immediately. A similar role is played by  $a(t)$  for buy orders. There is a strong coupling between  $b(t)$  and  $a(t)$ . Smith *et al.* (2003) observed how this nonlinear coupling makes modelling the LOB such a difficult problem.

### E. Quantifying patience

In a LOB, both "patient" and "impatient" market participants experience benefits and drawbacks for their actions. Patient market participants stand a chance of executing their trades at a better price than do impatient ones, but they also run the risk of their orders never being matched and of future adverse selection. Conversely, impatient traders never trade at prices better than  $b(t)$  and  $a(t)$ , but they do not face the inherent uncertainty associated with placing orders that do not match immediately. Foucault *et al.* (2005) conjectured that arbitrageurs, technical traders, and indexers were most likely to place impatient orders (due to the fast-paced nature of their work) and that portfolio managers were most likely to place patient orders (because their strategies are generally more focused on the long term). In reality, many market participants use a combination of both patient and impatient strategies in their interaction with the LOB, selecting their actions for each trade based on their individual needs at that time (Anand *et al.*, 2005). The bid-ask spread  $s(t)$  may be considered to be a measure of how highly the market values the immediacy and

certainty associated with market orders versus the waiting and uncertainty associated with limit orders.

Copeland and Galai (1983) noted that a limit order can be thought of as a derivative contract written to the whole market, via which the order's owner offers to buy or sell the specified quantity of the asset at the specified price to any market participant wishing to accept. For example, if market participant  $A$  submits a sell limit order  $x = (p_x, \omega_x)$ , this is equivalent to  $A$  offering the entire market a call option to buy  $\omega_x$  units of the asset at price  $p_x$ , so long as the order is active. Market participants offer such derivative contracts – i.e., submit limit orders – in the hope that they will be able to trade at better prices than if they simply submitted market orders. However, whether or not such a contract will be accepted by another market participant (i.e., whether or not the limit order will eventually become matched) is uncertain.

### F. Priority

As shown in Figure 1, it is possible to have active orders owned by different market participants at the same price at any given time. Much like priority is given to active orders with the best (i.e., highest buy or lowest sell) price, limit order markets also employ a priority system for active orders within each individual price level.

By far the most common priority mechanism currently used is *price-time*. That is, priority is first given to those active orders with the best price, and ties are broken by selecting the active order that was placed first among those. As Parlour (1998) highlighted, price-time priority is an effective way to encourage market participants to place limit orders in the LOB. Without a priority mechanism based on time, there is no incentive for a market participant to "show their hand" by submitting limit orders at any time before the desired price becomes the best price.

Another priority mechanism is *pro-rata*, which is commonly used in futures markets (Field and Large, 2008). Under this mechanism, when a tie needs to be broken, each active order participating in the tie-break receives a matching proportional to the fraction of the depth available that it represents at the best price. For example, if a buy market order of size  $3\sigma$  arrived at the LOB displayed in Figure 3,  $\sigma$  of it would match to active order 1 and  $2\sigma$  of it would match to active order 2, because they correspond to  $1/3$  and  $2/3$  of the depth available at  $a(t)$ , respectively.

Market participants trading in a pro-rata priority market are also faced with the substantial difficulty of optimally selecting limit order sizes, as posting limit orders with larger sizes than the quantity that is really desired for trade becomes a viable strategy to gain priority until a sufficiently large quantity has been matched, at which



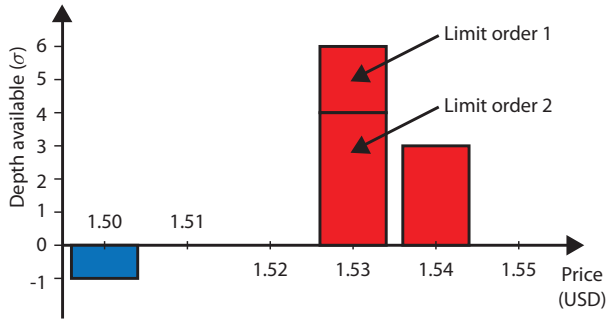


FIG. 3 Illustration of a LOB with pro-rata priority

time any remaining orders can be cancelled.

Another alternative priority mechanism is *price-size*, in which ties are broken by selecting the active order of the largest size among those limit orders with the best price. Until recently, no major exchanges used this priority mechanism, but in October 2010 the first price-size trading platform, NASDAQ OMX PSX, was launched (NASDAQ, 2010). Some exchanges allow market participants to specify a minimum match size when submitting orders. Other orders with a size smaller than this are not considered for matching against such orders. This may be considered to be a “pseudo-price-size” priority mechanism: small active orders are often bypassed, effectively giving higher priority to larger orders.

Different priority mechanisms encourage market participants to behave in different ways. Price-time priority encourages market participants to act quickly to ensure that their active orders sit at the front of the priority queue for their chosen price, and price-size and pro-rata priority reward traders for placing large limit orders and thus for providing greater liquidity to the market.

Because market participants’ behaviour is closely related to the priority mechanism used in the specific market, LOB models need to take priority mechanisms into account when considering order flow. Furthermore, in models that attempt to track specific orders, priority plays a pivotal role.

### G. Incomplete sampling and hidden liquidity

The LOB  $L(t)$  reflects only the subset of trading intentions that market participants have *announced* up to time  $t$ . However, the fact that a market participant hasn’t announced a desire to trade at a given price does not mean that they do now want to do so, as they may be keeping their intentions private from other market participants by only submitting orders when absolutely necessary. Unsurprisingly, quantifying market participants’ undisclosed trading intentions presents substantial difficulties when building models.

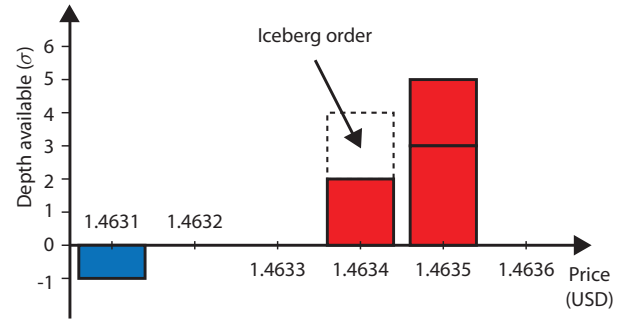


FIG. 4 Illustration of a LOB containing iceberg orders. Market participants would only see a depth available of  $n(1.4634, t) = 2\sigma$ , not  $4\sigma$ .

Bouchaud *et al.* (2009) highlighted that a typical snapshot of a LOB at a given time is very sparse, containing some active orders that are close to  $b(t)$  and  $a(t)$ , but also active orders that are very far from these prices. As described above, this cannot be interpreted as an indication that few people wish to trade at prices far from  $b(t)$  and  $a(t)$ , however; it is merely an indication that they have not announced any intention to do so. Indeed, some market participants choose not to submit limit orders at all.<sup>8</sup> These traders instead watch how the values of  $b(t)$  and  $a(t)$  evolve with time and place market orders when certain criteria are met.

Furthermore, many exchanges now allow market participants to conceal the extent of their intentions to trade, often at the cost of paying some penalty in terms of priority or price. As numerous authors (e.g., (Biais *et al.*, 1995; Bouchaud *et al.*, 2009; Hollifield *et al.*, 2004)) have highlighted, this poses significant problems when building models of LOBs.

A common example of how market participants can hide the extent of their intentions to trade is by using *iceberg orders* (also known as *hidden-size orders*). An iceberg order is a type of limit order that specifies not only a total size and price but also a *visible size*. Other market participants then only see the visible size. Rules regarding the treatment of the “hidden” quantity vary greatly from one exchange to another. In some cases, once a quantity of at least the visible size is matched to an incoming market order, another quantity equal to the visible size becomes visible, with a time priority position equal to that of a standard limit order placed at this time. This sort of iceberg order is similar to a market participant watching the market very carefully and submitting a new limit order at the same price and size at the exact moment that their previous limit order is matched to an

<sup>8</sup> Arbitrageurs are a key example of this. Their strategies depend on simultaneously buying and selling in an attempt to make instant profit. Limit orders are of little use to them because it is uncertain when (if ever) they will be matched.

incoming market order. A market participant acting in this way is sometimes deemed to be constructing a *synthetic iceberg order*. The only difference between a synthetic iceberg order and a genuine iceberg order occurs when a market order with a size larger than the (visible) depth available at the best price arrives. In this situation, the market order matches to any visible portions of active orders at the best price (according to the usual priority rules) and then a portion of any hidden depth available at this price. By contrast, if a market participant was submitting small but entirely visible duplicate limit orders immediately after their previous orders were matched, a large incoming market order would be matched only to the active orders that existed at that time, and the rest of the incoming market order would instead be matched to the active orders at the next best price. For example, if the sell order at \$1.4634 highlighted in Figure 4 were an iceberg order with a visible size of  $2\sigma$  and a hidden size of  $2\sigma$ , then an incoming buy market order of size  $4\sigma$  would have all  $4\sigma$  matched at \$1.4634, resulting in an ask-price of \$1.4635 immediately after the matching. However, if the highlighted sell order were a standard limit order of size  $2\sigma$ , whose owner was adopting the strategy of resubmitting a similar limit order of size  $2\sigma$  immediately after the original order was matched, then  $2\sigma$  of the incoming market order would be matched at \$1.4634 and  $2\sigma$  of the incoming market order would be matched at \$1.4635, followed by the new (duplicate) limit order repopulating the book at \$1.4634, resulting in an ask-price of \$1.4634 immediately after the submission of the limit order.

Some exchanges have an alternative structure for iceberg orders. Whenever a quantity equal to at least the visible size of an iceberg order is matched to an incoming market order, the rest of the order (i.e., the portion of the hidden component that is not also matched to the same incoming market order) is cancelled. In this way, market participants with an active iceberg order at a given price can match incoming market orders of a larger size than is initially apparent, without revealing to the market the true extent of his/her desire to trade (because only the visible portion of the order is displayed in the LOB), but otherwise the iceberg order behaves like any other order. This is the system currently used by the Reuters trading platform (Thomson-Reuters, 2011).

Some other trading platforms (e.g., Currenex and Hotspot FX (Gould *et al.*, 2011)) allow entirely hidden limit orders. These orders are given priority behind both entirely visible active orders at their price and the visible portion of iceberg orders at their price, but they give market participants the ability to submit limit orders without revealing any information whatsoever to the market.

Recently, there has also been an increase in the popularity of so-called *dark pools* (see, e.g., (Carrie, 2006; Hendershott and Jones, 2005)), particularly in equities trading. In a dark pool, no information about market participants' trading intentions is available to other

market participants. Rules governing matchings in dark pools vary greatly from one exchange to another (Mittal, 2008). Some dark pools are essentially LOBs in which all limit orders are entirely hidden; whereas other dark pools do not allow market participants to specify prices at all. Instead, they state only the size of their order and whether they wish to buy or sell. Orders are held in a time priority queue until they are matched to orders of the opposite type, and any trades that occur do so either at the mid-price  $m(t)$  of some other specified standard (i.e., non-dark) LOB in which the same asset is being traded, or at a price that is later negotiated by the two market participants involved.

Even in LOBs with no hidden liquidity, market participants are not always able to view the entire LOB in real time. In many exchanges, only limit orders that lie within a certain range of relative prices are displayed. On some electronic trading platforms, updates to  $L(t)$  are only transmitted by the exchange with a given frequency, meaning that all activity that has taken place since the most recent refresh signal is invisible to market participants. As discussed in Section III.A, market participants' actions have been found to vary with the amount of the LOB that they are able to view in real time (Boehmer *et al.*, 2005; Bortoli *et al.*, 2006).

## H. Volatility

Loosely speaking, volatility is a measure of the variability of returns of a traded asset (Barndorff-Nielsen and Shephard, 2010). There are many different measures of volatility commonly calculated from financial time series, and the exact form of volatility that is studied in a given situation depends heavily on the type of data that is available and the desired purpose of the calculation (Shephard, 2005). Even when their estimation is based on the same data, different measures of volatility sometimes exhibit different properties. For example, using data from a wide range of different markets, different measures of volatility have been found to follow different intra-day patterns; see (Cont *et al.*, 2011) and references therein. For this reason, many empirical studies report their results using more than one measure of volatility.

The volatility of an asset provides some indication of how risky investing in it is. All other things being equal, an asset with higher volatility should be expected to undergo larger price changes over a given time interval than an asset with lower volatility. For market participants who wish to carefully manage their risk exposure, volatility is a primary consideration when deciding which assets to invest in, and therefore forms the basis of optimal portfolio construction in the traditional economics literature.

As discussed in Section IV.F, links between volatility and several other important properties have been empirically observed in a wide range of LOBs.

## 1. Model-based estimates of volatility

A difficulty that arises when estimating any measure of volatility is that in a LOB, many market participants submit then immediately cancel buy (respectively, sell) limit orders within the spread. This can occur for a variety of reasons, but often it is the result of electronic trading algorithms searching for hidden liquidity. This causes  $b(t)$  (respectively,  $a(t)$ ) to fluctuate without any trades occurring, and may be considered *microstructure noise* rather than a meaningful change in price. One way to address this problem is to assume that the observed data is governed by a model from which an estimate of volatility that is absent of microstructure noise can be derived. The parameters of the model are then estimated from the data, and the volatility estimate is derived explicitly from the model. However, a drawback of this method is that the estimate of volatility that is obtained depends heavily on the model used, and models that poorly mimic some important aspect of the trading process may therefore output misleading estimates of volatility. For this reason, we restrict our attention to *model-free estimates of volatility*.

## 2. Model-free estimates of volatility

There is an extensive literature about using historical LOB data to perform model-free estimates of volatility (see, e.g., (Aït-Sahalia *et al.*, 2005; Andersen and Todorov, 2010; Bandi and Russell, 2006; Zhou, 1996)). In this section, we introduce three methods for performing such estimates.

**Definition.** Given the bid-price time series  $b(t_1), b(t_2), \dots, b(t_n)$  sampled at a regularly spaced set of times  $t_1, t_2, \dots, t_n$ , the bid-price realized volatility, denoted  $v_b(t_1, t_2, \dots, t_n)$ , is the standard deviation of the set of logarithmic returns  $\{r_b(t_i, t_{i+1}) \mid i = 1, 2, \dots, n-1\}$ . The ask-price realized volatility, denoted  $v_a(t_1, t_2, \dots, t_n)$ , and the mid-price realized volatility, denoted  $v_m(t_1, t_2, \dots, t_n)$ , are defined similarly.

Realized volatility is a useful measure for comparing the variability of return series that are sampled with the same fixed, specific frequency. For example, by using the daily closing mid-price of each of two stocks over the same year, a comparison of the realized volatility of each would provide insight into which stock's mid price varied more, day-on-day, in the given year. A similar comparison could be made between the mid price of the two stocks, recorded at the start of each second during a given trading day, to provide some insight into the relative size of the stocks' volatility during that day. However, realized volatility depends on the frequency at which the price series is sampled, so it is not appropriate to com-

pare the realized volatility of a once-daily price series for one stock to a once-hourly price series for another.

**Definition.** Given the bid-price time series  $b(t_1), b(t_2), \dots, b(t_n)$  sampled at the set of times  $t_1, t_2, \dots, t_n$  at which  $n$  consecutive sell market orders arrive, the bid-price realized volatility per trade, denoted  $V_b(t_1, t_2, \dots, t_n)$ , is the standard deviation of the set of logarithmic returns  $\{r_b(t_i, t_{i+1}) \mid i = 1, 2, \dots, n-1\}$ . The ask-price realized volatility per trade, denoted  $V_a(t_1, t_2, \dots, t_n)$ , is defined similarly, using  $n$  consecutive buy market order arrival times for  $t_1, t_2, \dots, t_n$ . The mid-price realized volatility per trade, denoted  $V_m(t_1, t_2, \dots, t_n)$ , is defined similarly, using  $n$  consecutive market order arrival times (irrespective of whether they are buy or sell market orders) for  $t_1, t_2, \dots, t_n$ .

Realized volatility per trade is a useful measure for comparing how prices vary on a trade-by-trade basis.

**Definition.** For a given trading day  $D$ , the bid-price intra-day volatility, denoted  $\rho_b(D)$ , is the logarithm of the ratio of the highest bid price during day  $D$  and the lowest bid price during day  $D$ :

$$\rho_b(D) = \log \left( \frac{\max_{t \in D} b(t)}{\min_{t \in D} b(t)} \right).$$

The ask-price intra-day volatility over trading day  $D$ , denoted  $\rho_a(D)$ , and the mid-price intra-day volatility over trading day  $D$ , denoted  $\rho_m(D)$ , are defined similarly.

Intra-day volatility is a useful measure of how likely very large price swings are in a given day. It is particularly important for *day traders*, who buy or sell assets in the market but unwind their positions before the end of each trading day. By knowing the intra-day volatility of an asset over a number of recent days, day traders can manage their exposure to the risk of a large price movement in a given day.

## 3. Time window effects

In order to estimate how the volatility of a given asset varies through time, it is common to use a *rolling time window* approach to estimate a *volatility series*. More precisely, the volatility of the asset is estimated over some given time window, then the time window is advanced by some predetermined amount. This process is repeated multiple times, so as to span the desired time range.

Figure 5 illustrates several points about using rolling time windows to estimate volatility series. First, the estimate of volatility is substantially larger whenever a spike in the original time series falls within the estimation time window. Using longer time windows causes such spikes to remain within them for a longer period of time, and therefore causes the volatility series estimate to be large

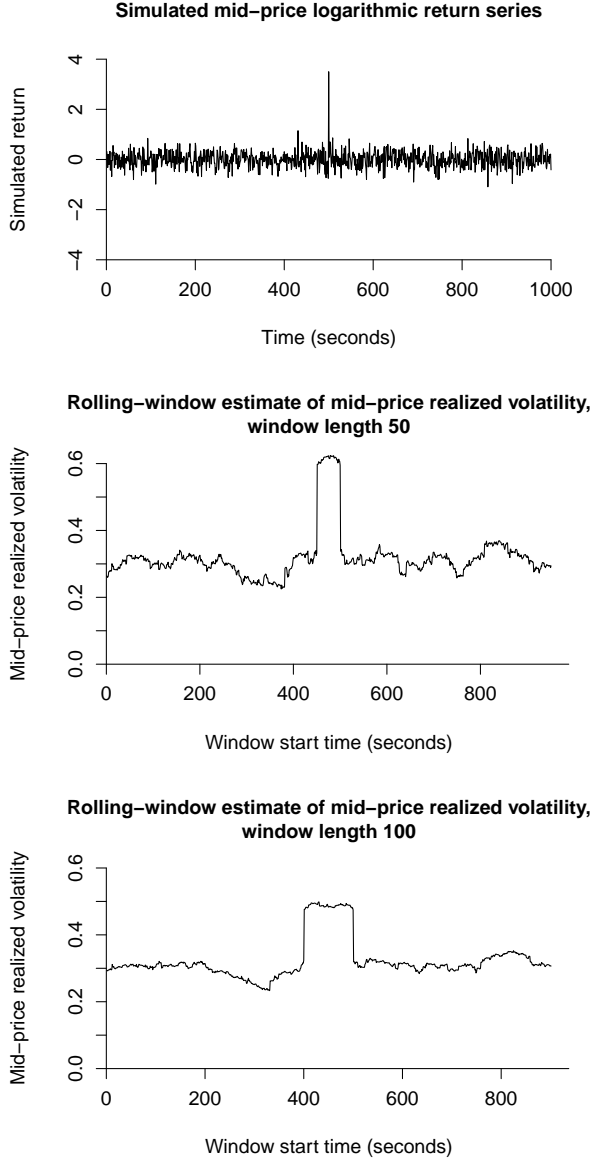


FIG. 5 Simulated mid-price logarithmic return series and two examples of corresponding mid-price realized volatility series

for longer. Second, longer time windows contain more data points, so the contribution to the volatility estimate made by any single point within the window is smaller. This means that a given spike in the original time series results in a smaller increase in the estimate of volatility when a longer time window is used. Third, longer time windows lead to smoother estimated volatility series, with or without spikes, because more observations are averaged within each window. For further discussion of these issues in a practical context, see Liu *et al.* (1999).

## I. Resolution parameters

To attract market participants, many new trading platforms offer smaller values of  $\sigma$  or  $\delta p$  than those offered by older trading platforms. Furthermore, values of  $\sigma$  and  $\delta p$  vary greatly from market to market. Expensive shares are normally traded with  $\sigma = 1$  share; cheaper shares are often traded with  $\sigma \gg 1$  share. In many foreign exchange markets, the currency pair XXX/YYY is traded with  $\sigma = 1$  million units of XXX, in others it is traded with as little as  $\sigma = 0.01$  units of XXX.<sup>9</sup> In equity markets,  $\delta p$  is often chosen to be around 0.01% of the mid-price  $m(t)$ , rounded to the nearest power of 10. For example, Apple Inc. stock's  $m(t)$  fluctuated in the range of \$150 to \$350 dollars during the year 2010, during which time it traded with  $\delta p = \$0.01$ . Common examples of  $\delta p$  in foreign exchange markets are around 0.001% of the mid-price  $m(t)$ , again rounded to the nearest power of 10. For example, on the electronic trading platform Hotspot FX,  $\delta p$  is \$0.00001 for GBP/USD trades (where  $m(t)$  fluctuated in the range \$1.40 to \$1.70 during 2010) and 0.001 yen for USD/JPY trades (where  $m(t)$  fluctuated in the range 80 yen to 100 yen during 2010).

It is a recurring theme in the empirical literature (see, e.g., (Biais *et al.*, 1995; Foucault *et al.*, 2005; Seppi, 1997; Smith *et al.*, 2003)) that a market's resolution parameters  $\sigma$  and  $\delta p$  greatly affect the way in which its market participants trade. Empirical regularities that are present in one market may, therefore, not be present in another, making the task of building a single universal LOB model very difficult.

## J. Bilateral trade agreements

On some exchanges, each market participant maintains a *trade agreement list* of other market participants with whom they are willing to trade. Under such a setup, a trade can only occur between market participants *A* and *B* if *A* appears on *B*'s trade agreement list, and vice-versa. The exchange shows each market participant a personalized version of the LOB that contains only the active orders owned by the market participants with whom it is possible for them to trade. When a market participant submits a market order, it will only match to an active order that has been displayed to that specific market participant, bypassing any higher priority active

<sup>9</sup> In foreign exchange markets, the XXX/YYY LOB is used to match exchanges of currency XXX to currency YYY. A price in the LOB XXX/YYY denotes how many units of currency YYY are being exchanged for a single unit of currency XXX. For example, a trade at the price \$1.52342 in the GBP/USD market would correspond to 1 GBP (i.e., pounds sterling) being exchanged for 1.52342 USD (i.e., US dollars).

orders from market participants with whom it is not possible for them to trade.

In exchanges operating such *bilateral trade agreements*, it is possible for a buy (respectively, sell) market order to bypass all active orders at the globally lowest (respectively, highest) price available in  $L(t)$ , and match to an active order with a strictly higher (respectively, lower) price. Furthermore, it is possible for  $L(t)$  to simultaneously contain both an active buy order  $x = (p_x, \omega_x)$  owned by a market participant  $A$  and an active sell order  $y = (p_y, \omega_y)$  owned by a market participant  $B$ , with  $p_y \leq p_x$ , without a matching occurring, if  $A$  and  $B$  do not have a bilateral trade agreement. Therefore, it is possible for  $s(t) \leq 0$ .

All of these factors make modelling of specific matchings and of the evolution of  $L(t)$  a very difficult task in LOBs that operate with bilateral trade agreements. Gould *et al.* (2011) presents a full discussion of these issues, so we do not consider such LOBs further here.

### K. Opening and closing auctions

Many exchanges suspend standard limit order trading at the beginning and end of the trading day, instead using an auction system to match orders. For example, the LSE's flagship order book SETS (SETS, 2011) has three distinct trading phases in each trading day. Between 08:00 and 16:30, the standard LOB mechanism is used, in a period known as "continuous trading". Between 07:50 and 08:00, a 10 minute "opening auction" takes place, and between 16:30 and 16:35, a 5 minute "closing auction" takes place. Both auctions use the same rules. During these periods, all market participants can view and place orders as usual, but no orders are matched. Due to the absence of matchings, the highest price among buy orders is allowed to (and often does) exceed the lowest price among sell orders. All orders are stored until the opening auction period ends. At this time, for each price  $p$  at which there is non-zero depth available, the trade matching algorithm calculates the number  $C_p$  of trades that could occur by matching buy orders with a price greater than or equal to  $p$  to sell orders with a price less than or equal to  $p$ . The *uncrossing price*:

$$\hat{p} = \arg \max_p C_p \quad (3)$$

is calculated, and all relevant matchings occur at this price. Crucially, and in contrast to standard limit order trading, all trades take place at the same uncrossing price  $\hat{p}$ , irrespective of what price the original orders specified. There are specific rules for choosing  $\hat{p}$  if multiple values of  $p$  solve equation (3). Once the uncrossing price  $\hat{p}$  has been determined, if there is a smaller depth available for sale than there is for purchase (or vice versa), ties are broken using time priority.

Throughout the opening auction, all market participants can see what the values of  $\hat{p}$  and  $A_{\hat{p}}$  would be if the auction were to end at that moment. The purpose of the opening auction is to allow all market participants to observe the "discovery" of the price without any matchings taking place until the process is complete. Such a price discovery process is common to many markets.<sup>10</sup>

### L. Power laws

Several LOB properties have been reported to have *power-law tails*:

**Definition.** A random variable  $Z$  is said to have a power-law tail with exponent  $\alpha$  if, in the limit  $z \rightarrow \infty$ , there exists some  $\alpha \in \mathbb{R}$  such that its probability density function  $f_Z(z)$  decays like  $z^{-\alpha}$ ; i.e.,  $f_Z(z) \sim O(z^{-\alpha})$ .

If there exists some finite  $z_{\min} > 0$  such that  $f_Z(z)$  is proportional to  $z^{-\alpha}$  for all  $z \geq z_{\min}$ , then clearly  $Z$  has a power-law tail. Although this is far from being the only possible probability density function to describe a power-law tail, the existence of such a  $z_{\min}$  allows simple closed-form expressions to be derived for when  $Z \geq z_{\min}$  (Clauset *et al.*, 2009). In particular, if  $Z$  is a continuous random variable for which  $f_Z(z) = kz^{-\alpha}$  for all  $z \geq z_{\min}$ , with  $\alpha > 1$ , then the probability density function  $f_Z(z|Z \geq z_{\min})$  describing the distribution of  $Z$  restricted to the region  $Z \geq z_{\min}$  can be calculated explicitly by noticing that the integral of  $f_Z(z|Z \geq z_{\min})$  over its domain must equal 1, i.e.,

$$\int_{z_{\min}}^{\infty} f_Z(z|Z \geq z_{\min}) = \int_{z_{\min}}^{\infty} kz^{-\alpha} = 1. \quad (4)$$

Therefore:

$$k = \frac{\alpha - 1}{z_{\min}^{-\alpha+1}}. \quad (5)$$

Yielding:

$$f_Z(z|Z \geq z_{\min}) = \frac{\alpha - 1}{z_{\min}} \left( \frac{z}{z_{\min}} \right)^{-\alpha}, \quad z \geq z_{\min}. \quad (6)$$

When attempting to fit power-law tails to empirical observations, it is often assumed that  $f_Z(z|Z \geq z_{\min})$  takes

<sup>10</sup> Biais *et al.* (1999) performed a formal hypothesis test on price discovery data from the Paris Bourse. Working at the 2.5% level, they did not reject the null hypothesis that market participants "learned" (i.e., that their conditional expectations approached the market value of the asset) during the final 9 minutes of the price discovery process. However, they found that during the early part of the price discovery process, market participants' actions were not significantly different from noise.

this functional form, in order to make use of such closed-form expressions in the inference process. Under this assumption, Clauset *et al.* (2009) provides a comprehensive algorithm for consistent estimation of  $\alpha$  and  $z_{\min}$  directly from empirical data via maximum likelihood techniques, and for formally testing the hypothesis that data really does follow a power law. Several other consistent estimation procedures also exist (e.g., (Hill, 1975; Mu *et al.*, 2009)), but no single estimator has emerged as the best to adopt in all empirical analysis, where the number of samples is inevitably finite. For this reason, some empirical studies report the estimates made by several different estimators, then draw inference about  $\alpha$  based on the whole set of results. However, as Mu *et al.* (2009) highlighted, in many situations different estimators produce vastly different estimates of  $\alpha$ , making such inference difficult.

Despite their prominence throughout the scientific literature, doubt has recently been cast over the validity of many reported power-laws (Clauset *et al.*, 2009; Stumpf and Porter, 2012). When assessing whether empirical data might follow a power law, it is crucial to perform goodness-of-fit and likelihood ratio tests, in order to provide a formal assessment of the fit and to compare it against other candidate distributions. Clauset *et al.* (2009) noted that the evidence supporting many reported power-laws in empirical studies consisted of little more than the observation of an approximately straight line on a log-log plot of the data. Furthermore, they demonstrated that estimating  $\alpha$  by performing a linear least-squares regression on such a log-log plot led to significant systematic errors, as several of the basic assumptions required to apply least-squares regression did not apply. However, many of the power laws reported in empirical studies of LOBs are not accompanied by any such goodness-of-fit or likelihood ratio tests, and many authors explicitly state that their estimation of the power-law exponent was performed via linear least-squares regression on a log-log plot. Furthermore, when working with power laws, estimators and test statistics take different functional forms for discrete data than they do for continuous data (Clauset *et al.*, 2009). Using the estimators derived in the continuous case on discrete data leads to substantial systematic errors, yet many empirical publications do precisely this.

## M. Long-Range Correlations

As discussed in Section IV.H, a wide variety of time series related to LOBs have been reported to exhibit long memory: i.e., long-range autocorrelations. If the exact correlation structure is known (or well estimated) and a long history of the series has been observed, such long-range autocorrelations provide substantial benefits when forecasting future values of a time series, as the autocor-

relation structure makes future values more predictable. However, estimation of the properties of such time series is laden with technical and practical difficulties that are rarely acknowledged in the empirical literature. Beran (1994) surveyed a wide range of technical results regarding the convergence of estimators for time series with long memory; here we focus on the practical challenges of estimating long-range autocorrelations, paying careful attention to the associated difficulties.

Throughout this section, we denote by  $\mathbf{X}$  a second-order stationary time series  $\mathbf{X} = X(t_1), X(t_2), \dots, X(t_k)$ ,<sup>11</sup> and define autocorrelation as follows:

**Definition.** The lag- $l$  autocorrelation of a time series  $\mathbf{X}$  is given by

$$A_{\mathbf{X}}(l) = \frac{1}{k-l} \sum_{i=1}^{k-l} (X(t_i) - \langle \mathbf{X} \rangle) (X(t_{i+l}) - \langle \mathbf{X} \rangle),$$

where  $\langle \mathbf{X} \rangle = \frac{1}{k} \sum_{i=1}^k X(t_i)$  is the time-averaged value.

When considered as a function of  $l$ ,  $A_{\mathbf{X}}$  is called the *autocorrelation function*. For clarity, we initially restrict our attention to processes with positive long-range autocorrelations. Time series with negative long-range autocorrelations certainly exist, but the discussion of their properties becomes more cluttered due to  $\pm$  sign considerations. Negative long-range autocorrelations are reintroduced from Section III.M.3 onwards.

### 1. Long- and short-memory processes

**Definition.** The time series  $\mathbf{X}$  is said to exhibit short memory if, in the limit  $l \rightarrow \infty$ , the autocorrelation function  $A_{\mathbf{X}}$  decays exponentially in  $l$ ; i.e.,  $A_{\mathbf{X}}(l) \sim O(e^{-l/\tau})$ ,  $l \rightarrow \infty$ .

For such short-memory processes,  $\tau$  is an indication of the number of time steps over which  $\mathbf{X}$  is appreciably autocorrelated. An example of a short-memory process is an autoregressive process (Hamilton, 1994).

**Definition.** The time series  $\mathbf{X}$  is said to exhibit long memory if, in the limit  $l \rightarrow \infty$ , the autocorrelation function  $A_{\mathbf{X}}$  decays like a power law in  $l$ ; i.e.,  $A_{\mathbf{X}}(l) \sim O(l^{-\alpha})$ ,  $l \rightarrow \infty$ , where  $0 < \alpha < 1$ .

The exponent  $\alpha$  describes the strength of the long memory: the smaller the value of  $\alpha$ , the longer the memory of the process.

<sup>11</sup> A time series is second-order stationary if its first and second moments are finite and do not vary with time. For a discussion of issues regarding stationarity in financial time series, see Taylor (2008).

## 2. Practical considerations

The above definitions present difficulties in practice. First, both definitions deal only with asymptotic behaviour, and do not specify autocorrelations at any finite  $l$ . Clearly, any empirically-observed time series is finite, so judgement must be made as to when, or, indeed, whether, it is appropriate to judge the autocorrelation function as approaching its asymptotic behaviour. Second, the definitions deal only with rates of convergence. Autocorrelations may themselves be arbitrarily small, making their estimation very difficult. Third, for a long-memory process, values from the distant past can have a statistically significant impact on values in the present. Therefore, the estimation of parameters and their confidence intervals is a difficult task, as the long-range autocorrelation causes the effective sample size to be far smaller than it might initially seem (Farmer and Lillo, 2004).

If samples are independent and identically distributed, the variance of many estimators scales with the sample size  $k$  at a rate  $O(k^{-\frac{1}{2}})$ . However, in the presence of long memory, the variance of many such estimators scales with the sample size  $k$  at a rate slower than  $O(k^{-\frac{1}{2}})$ ; meaning that a very large sample is required in order to make good estimates. In many cases, if it is erroneously assumed that samples from a long-memory process are actually uncorrelated, the probability that a standard confidence interval of an estimate contains the true value of the parameter tends to 0 as  $k \rightarrow \infty$  (Beran, 1994). Therefore, modifications to such calculations must be made in order to achieve sensible confidence intervals that take the long-range autocorrelations into account.

For these reasons, along with the potential existence of noise or trends in  $\mathbf{X}$ , direct estimation of  $\alpha$  from the autocorrelation function  $A_{\mathbf{X}}(l)$  often produces very poor results (Lillo and Farmer, 2004). Although log-log plots of  $A_{\mathbf{X}}$  can be useful as a preliminary visual tool to informally assess whether  $\mathbf{X}$  might have long memory, such plots are of little use beyond this. Instead, a variety of techniques for obtaining better estimates of the strength of long memory have been developed, as we now discuss.

## 3. The Hurst exponent

Given  $\mathbf{X}$ , define the new time series  $\mathbf{Y}$  as the partial sums of  $\mathbf{X}$ :

$$Y(t_i) = \sum_{j=1}^i X(t_j), i = 1, \dots, k.$$

In this way,  $\mathbf{Y}$  can be thought of as the random walk whose jump at time  $t_i$  is given by  $X(t_i)$ . It can be shown that the standard deviation of  $Y(t_{i+l}) - Y(t_i)$

scales asymptotically like  $l^H$ , for some  $H$ . This follows from the fact that

$$Y(t_{i+l}) - Y(t_i) = \sum_{j=1}^{i+l} X(t_j) - \sum_{j=1}^i X(t_j) = \sum_{j=i+1}^{i+l} X(t_j), \quad (7)$$

then using the asymptotic properties of the sums of such  $X(t_i)$ .  $H$  is known as the *Hurst exponent* of  $\mathbf{X}$ , and is related to  $\alpha$ , discussed above, according to

$$H = 1 - \frac{\alpha}{2}. \quad (8)$$

Therefore,  $H$  also provides information about the strength of long memory in a time series. If  $\mathbf{X}$  has a Hurst exponent of  $\frac{1}{2}$ , then it does not display long memory. If  $\mathbf{X}$  has a Hurst exponent  $\frac{1}{2} < H < 1$ , then  $\mathbf{X}$  has long memory with positive long-range autocorrelations. It is a recurring mistake in the literature that if  $\mathbf{X}$  has Hurst exponent  $\frac{1}{2} < H < 1$ , its unconditional distribution  $f(X(t_i))$  must exhibit heavy tails. However, Preis *et al.* (2006, 2007) showed that such an implication does not hold in general.

Time series with negative long-range autocorrelations can also be classified using their Hurst exponent. In particular, if  $\mathbf{X}$  has Hurst exponent  $0 < H < \frac{1}{2}$ , then  $\mathbf{X}$  has long memory with negative long-range autocorrelations.

## 4. Estimators of $H$

Under some assumptions on  $\mathbf{X}$ ,<sup>12</sup> there are several estimators of  $H$  that are unbiased in the limit of infinite sample size  $k$  (Taqqu *et al.*, 1995), and that are more robust against noise in the underlying time series than is estimation of the  $\alpha$ . However, the performance of such estimators on empirical data, which is finite in size and may not conform to these assumptions, varies considerably. Rea *et al.* (2009) reviewed both the asymptotic properties and finite-sample performance of several Hurst-exponent estimators. Different estimators are more common in different disciplines, although such choices tend to be for historical reasons, rather than based on performance. Some of the most commonly-used estimators are:

- The *R/S statistic* and *modified R/S statistic*, which examine the scaling of the difference between the maximum and minimum departures of  $\mathbf{Y}$  from a random walk in which all jump sizes are equal to the mean jump size. However, short-range autocorrelations are known to affect the performance of the R/S statistic when applied to finite data sets,

<sup>12</sup> Most commonly, it is assumed that  $\mathbf{X}$  is a fractional Brownian motion (i.e., a self-similar process with stationary Gaussian increments (Beran, 1994; Robinson, 2003)).

and the modified R/S statistic often fails to detect long memory when in fact it is known to be present (Lo, 1989; Teverovsky *et al.*, 1999).

- *Log-periodogram regression*, which estimates  $H$  via ordinary least squares regression in the spectral domain, rather than directly from the autocorrelation function (Geweke and Porter-Hudak, 1983).
- *Order- $m$  detrended fluctuation analysis (DFAm)*, which examines how the standard deviation of terms of the form given in Equation (7) scales with  $l$ , but after first removing any local polynomial trends of order less than or equal to  $m$  (Kantelhardt *et al.*, 2001; Peng *et al.*, 1994). DFAm may be used on a time series that includes polynomial nonstationarities, and has been found to provide consistent estimates of  $H$  for a very general class of time series  $\mathbf{X}$  (La Spada and Lillo, 2011). Furthermore, Xu *et al.* (2005) found DFAm to produce estimates with smaller variance and smaller bias than those of several other estimators when applied to finite datasets. DFAm is also a useful tool for studying short-memory processes, as it is able to provide an estimate of  $\tau$  for such a process (Kantelhardt *et al.*, 2001; Peng *et al.*, 1994).

Much like with the estimation of power laws discussed in Section III.L, no single estimator has emerged as the best to adopt in all situations, so it has become common to report the estimates made by several estimators and to draw inference about  $H$  based on these results (Taqqu *et al.*, 1995). However, DFAm clearly offers substantial practical benefits over the other estimators discussed here.

#### IV. EMPIRICAL OBSERVATIONS IN LIMIT ORDER MARKETS

A wide range of LOB features have been studied in the empirical literature, often with conflicting conclusions. There are many possible reasons for such disagreements, including different markets operating differently (perhaps for cultural reasons or simply because trade matching algorithms operate differently), different asset classes being traded on different exchanges, differing levels of liquidity in different markets, and different researchers having access to different quality data. Furthermore, different authors have studied data from different years, and as market participants' trading strategies have evolved over time, so too have the statistical hallmarks they create. This is a particularly important consideration in recent years, as electronic trading algorithms have come to play an increasingly prominent role in markets and have caused an increase in both competition and trading volumes.

In order to aid comparisons between different studies, we present in Appendix A a description of all the empirical studies discussed in this survey that focus particularly on LOBs. We include information regarding the date range, source, and type of data that these studies were based on. As is clear from the table, a large number of aspects related to order placement have been studied (including the distribution of relative prices for newly arriving orders (Bouchaud *et al.*, 2002; Gu *et al.*, 2008b; Hollifield *et al.*, 2004; Potters and Bouchaud, 2003; Zovko and Farmer, 2002); the distribution of sizes for newly arriving limit and market orders (Bouchaud *et al.*, 2002; Challet and Stinchcombe, 2001; Gopikrishnan *et al.*, 2000; Maslov and Mills, 2001; Mu *et al.*, 2009); the arrival rate of orders (Biais *et al.*, 1995; Bouchaud *et al.*, 2002; Challet and Stinchcombe, 2001; Cont *et al.*, 2010; Maskawa, 2007; Mike and Farmer, 2008; Rinaldo, 2004)); the depth profile (Biais *et al.*, 1995; Bouchaud *et al.*, 2002; Gu *et al.*, 2008c; Hollifield *et al.*, 2004; Potters and Bouchaud, 2003; Roşu, 2009); order cancellation rates (Cao *et al.*, 2008; Challet and Stinchcombe, 2001; Hasbrouck and Saar, 2002; Potters and Bouchaud, 2003); and price changes (Cont *et al.*, 2011; Eisler *et al.*, 2010; Gu *et al.*, 2008a; Plerou and Stanley, 2008; Zhou, 2012). We now discuss the main findings of these empirical studies in more detail, and examine how such findings can deepen understanding about certain aspects of limit order trading. We also discuss a selection of *stylized facts* that have consistently emerged from several different empirical examinations of LOB data. As we highlight in Section V, such stylized facts can be a valuable tool for assessing LOB models, as a model's failure to reproduce the stylized facts is an indication that it poorly mimics some aspect of limit order trading.

##### A. Order size

When submitting a new order, a market participant must select its size. Given the heterogeneous motivations for trade that exist within a single market, it is unsurprising that there is a substantial variation in the size of incoming orders; yet a number of regularities have also been observed in empirical data.

For equities traded on the Paris Bourse, the distribution of  $\log(|\omega_x|)$  was reported to be approximately uniform for incoming limit orders with  $10 < |\omega_x| < 50000$  (Bouchaud *et al.*, 2002). For two stocks traded on NASDAQ, independent fits to the distribution of incoming limit order sizes  $|\omega_x|$  were made for different relative prices (Maslov and Mills, 2001). Power-law and log-normal distributions were reported. The mean reported power-law exponent was  $1 \pm 0.3$  (i.e., with standard deviation 0.3). However, the quality of the power-law fits was deemed to be weak; and the log-normal fits were deemed to be applicable over a wider range of limit or-



der sizes than the power-law fits (although the authors stated no precise range of applicability for either). For four stocks on the Island ECN, incoming limit order sizes  $|\omega_x|$  were reported to cluster strongly at “round number” amounts, such as 10, 100, and 1000 (Challet and Stinchcombe, 2001). A similar “round number” preference was observed for market orders on the Shenzhen Stock Exchange (Mu *et al.*, 2009). The authors also studied the distribution of total trade sizes when aggregated over a variety of time windows, and found it to exhibit a power-law tail. Different power-law exponent estimators were found to produce different estimates of the tail exponent, but overall the authors judged the tail exponent to be larger than 2. Similar power-law fits were also reported on NASDAQ (Maslov and Mills, 2001). Studying 5 days of data covering 3 equities altogether; the mean reported power-law exponent was  $1.4 \pm 0.1$ . Although the authors did not state a range of sizes over which their reported power-law distributions applied, Figure 1 in the publication (Maslov and Mills, 2001) suggests an approximate range of 200 to 5000. Power-law fits were also reported for the distribution of trade sizes in a study of the 1000 largest equities in the USA (Gopikrishnan *et al.*, 2000). The mean reported power-law exponent was  $1.53 \pm 0.07$ . However, Bouchaud *et al.* (2009) noted that the data studied by Gopikrishnan *et al.* contained information about trades that were privately arranged to occur “off-book”, and therefore were not conducted via the LOB. They conjectured that as larger traders were more likely to be arranged off-book, Gopikrishnan *et al.* had actually overestimated the frequency with which very large orders occurred in the LOB.

Buy (respectively, sell) market orders with a size  $|\omega_x| > n(a(t), t)$  (respectively,  $\omega_x > |n(b(t), t)|$ ) – i.e., those that “walk up the book” – were found to account for only 0.1% of submitted market orders on the Stockholm Stock Exchange (Hollifield *et al.*, 2004). Therefore, the vast majority of submitted buy (respectively, sell) market orders were found to match only to limit orders at  $a(t)$  (respectively,  $b(t)$ ), not at prices deeper into the LOB.

## B. Relative price

Along with its size, the other decision that a market participant must make when submitting an order is its price. As discussed in Section II.A, regularities in price series are best investigated via the use of relative pricing, as  $b(t)$  and  $a(t)$  themselves evolve through time. Unlike the distribution of order size, where different markets seem to exhibit different functional forms, the distribution of relative prices appears to exhibit a power-law behaviour in all studied markets. This shows that some market participants place limit orders deep into the LOB, which may suggest that they hold an optimistic belief that large price swings might occur.

Such a power law has been reported for the distribution of relative prices for orders that arrive with a non-negative relative price on the Paris Bourse (Bouchaud *et al.*, 2002), NASDAQ (Potters and Bouchaud, 2003), the LSE (Maskawa, 2007; Zovko and Farmer, 2002), and the Shenzhen Stock Exchange (Gu *et al.*, 2008b); however, different values of the exponent in the power law have been observed in different markets. A value of approximately 0.6 was reported to fit the distribution of relative prices from  $\delta p$  to over  $100\delta p$  (even up to  $1000\delta p$  for some stocks) on the Paris Bourse, for buy orders and sell orders alike. The power-law exponents, as well as the ranges of relative prices over which the distribution was reported to follow a power law, were found to vary across the stocks studied on NASDAQ (Potters and Bouchaud, 2003). For both buy and sell orders, the value of the power-law exponent was reported to be approximately 1.5 for relative prices between  $10\delta p$  and  $2000\delta p$  on the LSE. A power law with an exponent of  $1.72 \pm 0.03$  for buy orders and  $1.15 \pm 0.02$  for sell orders, was fitted to the distribution of non-negative relative prices<sup>13</sup> from the aggregated dataset of all 23 studied stocks on the Shenzhen Stock Exchange, however the exact matching rules on the Shenzhen Stock Exchange prevent large price changes from occurring within a single day, so some care must be taken when comparing these values to the ones from other markets. An asymmetry between buy orders and sell orders was reported for the Shenzhen Stock Exchange, but not for the other markets.

A power law was also reported for the distribution of relative prices for orders that arrive with a negative relative price on the Shenzhen Stock Exchange. The exponent was found to be  $1.66 \pm 0.07$  for buy orders and  $1.80 \pm 0.07$  for sell orders.

The distribution of all (i.e., both non-negative and negative) relative prices for the Astrazeneca stock on the LSE (Mike and Farmer, 2008) was found to follow a Student’s  $t$  distribution with 1.3 degrees of freedom. Such a distribution has power-law tails, with an exponent of 2.3. The mode of the distribution was found to occur at a relative price of 0 (i.e., at  $b(t)$  or  $a(t)$ ). The maximum arrival rate was also observed to occur at a relative price of 0 on the Shenzhen Stock Exchange (Gu *et al.*, 2008b), the Paris Bourse (Biais *et al.*, 1995; Bouchaud *et al.*, 2002), NASDAQ (Challet and Stinchcombe, 2001), and the LSE (Mike and Farmer, 2008), although on the Tokyo Stock Exchange the maximum was found to occur inside the spread (Cont *et al.*, 2010).

<sup>13</sup> Notice that the notation used by Gu *et al.* (2008b) assigns the opposite signs when measuring relative price than those used here.

### C. Limit order cancellations

Along with the submission of orders, cancellations play a major role in the evolution of  $L(t)$ . A limit order that was attractive to its owner at the time of submission may, for a variety of reasons, become unattractive at a later time, and therefore require cancellation to avoid an undesired matching. In a detailed study of activity on the Australian Stock Exchange, Cao *et al.* (2008) found that market participants viewed cancellations (and amendments, which are reported in most data sets as a cancellation followed by a new limit order submission) as an integral part of their strategy when dealing with limit order markets. Furthermore, cancellations play a central role for many electronic trading algorithms, which often use limit order submissions that are almost immediately followed by cancellations in an attempt to detect hidden liquidity (Hendershott *et al.*, 2011).

Approximately 25% of limit orders placed on the Island Electronic Communications Network (ECN) during the fourth quarter of 1999 were cancelled within two seconds of being placed, and 40% were cancelled within 10 seconds of being placed (Hasbrouck and Saar, 2002). Moreover, the observed rate of cancellations was found to decrease monotonically away from  $b(t)$  and  $a(t)$  (Potters and Bouchaud, 2003). A further study reported that as many as 80% of the limit orders for a selection of equities on Island ECN (namely, Cisco, Dell, Microsoft, and Worldcom) ended in cancellation rather than matching (Challet and Stinchcombe, 2001). The distribution of the length of time between placement and cancellation for cancelled orders was found to exhibit large spikes at 90, 180, 300, and 600 seconds. The authors proposed that such spikes could result from an automatic mechanism for cancelling active orders after a specified time.

### D. Mean relative depth profile

Despite their different resolution parameters (see Section II.A) and the different prices at which trades occur in them, a number of qualitative regularities have been empirically observed in mean relative depth profiles from a wide range of markets.

No significant difference has been detected between the mean bid-side and the mean ask-side relative depth profiles on the Paris Bourse (Biais *et al.*, 1995; Bouchaud *et al.*, 2002), NASDAQ (Potters and Bouchaud, 2003), and for Standard and Poor's Depositary Receipts (SPY)<sup>14</sup> (Potters and Bouchaud, 2003). Such symmetry was not reported on the Shenzhen Stock Exchange

(Gu *et al.*, 2008c), however this is unsurprising considering that this market has an additional rule restricting the size of price movements on any given day, which imposes an asymmetric restriction on the range of relative prices that may be selected after a price move has occurred.

Mean relative depth profiles have also been found to exhibit a “hump” shape<sup>15</sup> in a wide range of markets, including the Paris Bourse (Bouchaud *et al.*, 2002), NASDAQ (Potters and Bouchaud, 2003), the Stockholm Stock Exchange (Hollifield *et al.*, 2004), and the Shenzhen Stock Exchange (Gu *et al.*, 2008c). The maximal mean depth available for SPY was found to occur at  $b(t)$  and  $a(t)$ , which can also be considered as a hump with its maximum at a relative price of 0 (Potters and Bouchaud, 2003).

The exact location of the hump has been found to vary between different markets, although this is unsurprising given their different resolution parameters. The effect of changing resolution parameters is twofold: First, so long as  $\delta p$  is sufficiently small in a given market, further reduction in  $\delta p$  would cause the hump would reside at a larger number of ticks (i.e., multiples of  $\delta p$ ) from the best price than it did before, but only because the measurement scale had changed. Second, the resolution parameters might themselves affect the way in which market participants behave, as discussed in Section III.I.

There might, of course, also be strategic reasons that the hump occurs in different locations for different markets. For example, in markets where large price changes are relatively common, more market participants may choose to submit limit orders with larger relative prices than in those where such changes are rare, thereby increasing the relative price at which the hump resides. Roşu (2009) conjectured that a hump would exist in all markets where large market orders were sufficiently likely; this represents a trade-off between the optimism that a limit order placed away from  $b(t)$  or  $a(t)$  might be matched to a market order (at a significant profit for the limit order holder) and the pessimism that placing limit orders too far away from the current bid/ask might be a waste of time, as they might never be matched.

### E. Volatility

For the companies that make up the FTSE 100 (Zumbach, 2004), realized mid-price volatility (using 15 minute time windows) was found to be independent of the size of the market capitalization of the company, whereas mid-price volatility per trade was found to be lower for the companies with larger market capitalization.

<sup>14</sup> SPY is an exchange traded fund that allows market participants to effectively buy and sell shares in all of the 500 largest stocks traded in the USA.

<sup>15</sup> More precisely, the absolute value of the mean depth available has been found to increase monotonically over the first few relative prices, then to decrease monotonically beyond this, thereby creating a “hump” in the mean relative depth profile.

In foreign exchange markets (Zhou, 1996), hourly mid-price realized volatility was found to follow different intra-day patterns for different currency pairs. The authors conjectured that this was because different currencies became more heavily traded at different times of day, due to the different time zones. For all currency pairs studied, daily mid-price realized volatility was found to be higher later in the trading week.

On the Swiss Stock Exchange (Ranaldo, 2004), spread and volatility were found to follow strong intra-day patterns.

## F. Conditional event frequencies

The properties discussed so far in this section have all been calculated unconditionally, without reference to other variables. However, several factors influence how market participants interact with LOBs, so it is reasonable to believe that a deeper understanding might be achieved by studying not only unconditional frequencies, but also the frequencies of those events given that some other condition was satisfied. However, the study of such conditional event frequencies in LOBs is far from straightforward, for two main reasons:

1. The state space is very large. Deciding which of the enormous number of possible events or order book states to condition on is very difficult (Parlour and Seppi, 2008).
2. There is a small *latency* between the time that a market participant sends an instruction to submit or cancel an order and the time that the exchange server receives the instruction. Furthermore, as refresh signals are only transmitted by the exchange server at discrete time intervals, market participants cannot be certain that the LOB they observe via their trading platform is a perfect representation of the actual LOB at that instant in time. Therefore, conditioning on the “most recent” event is problematic, as the most recent event recorded by the exchange (and thus appearing in the market data) may not be the most recent event that a given market participant observed via the trading platform.

Nevertheless, several studies of conditional event frequencies in LOBs have identified interesting behaviour in empirical data. In this section, we review the key findings from several such publications, highlighting both the similarities and differences that have emerged across different markets. Most such studies have used older LOB data, often dating back 10 years or more. While this does help alleviate the difficulties with latency outlined above (as the volume of order flows in LOBs was much lower in the past than it is today, so the mean inter-arrival times

between successive events were substantially longer than the latency times), it also inevitably raises the question of how representative such findings are of limit order trading today.

### 1. Order size

A simple example of conditional structure is the relationship between  $|\omega_x|$  and the relative price  $\Delta_x$  of orders, reported on the Paris Bourse by Bouchaud *et al.* (2002). Initially, the unconditional distribution of  $|\omega_x|$  was estimated by examining the size of all orders that arrived, as discussed in Section IV.A. However, when the authors partitioned incoming orders according to  $\Delta_x$  and fitted a separate distribution for each value of  $\Delta_x$ , substantial variation was found between the fitted distributions. In particular, those orders with larger relative price were found to have a smaller  $|\omega_x|$  on average. A similar observation was made for limit orders on NASDAQ (Maslov and Mills, 2001).

### 2. Relative price

On the Paris Bourse (Biais *et al.*, 1995) and the Australian Stock Exchange (Cao *et al.*, 2008; Hall and Hautsch, 2006), market participants were found to place more orders with a relative price  $-s(t) < \Delta_x < 0$  (i.e., limit orders falling inside the spread) at times when  $s(t)$  was larger than its median value. Similarly, on the NYSE (Ellul *et al.*, 2003) the percentage of incoming orders that arrived with a relative price  $\Delta_x > -s(t)$  (i.e., were limit orders) was found to increase as  $s(t)$  increased, and was found to decrease when  $s(t)$  decreased. Biais *et al.* (1995) argued that when  $s(t)$  was small, it was less expensive for market participants to demand immediate liquidity, so more market orders were placed. However, it is also possible to explain such an observation via a zero-intelligence approach. If limit order prices are chosen uniformly at random over some fixed price interval, then it is more likely that an incoming limit order price resides in the interval  $(b(t), a(t))$  when the interval is wider.

The percentage of buy (respectively, sell) limit orders that arrived with a relative price  $-s(t) < \Delta_x < 0$  on the Paris Bourse (Biais *et al.*, 1995) was found to be higher at times when  $|n(b(t), t)|$  (respectively,  $n(a(t), t)$ ) was larger. This was conjectured to be evidence of market participants competing for priority, as the only way to gain higher priority than the active orders in the (already long) queue in this situation is to submit an order with a better price. Furthermore, on the NYSE (Ellul *et al.*, 2003) the arrival rate of buy (respectively, sell) limit orders with a relative price of  $-s(t) < \Delta_x < 0$  was found to increase as the total depth available on the bid (respectively, ask) side of the LOB increased, as

was the rate of buy (respectively, sell) market orders. Similarly, on the Australian Stock exchange (Hall and Hautsch, 2006), the percentage of buy (respectively, sell) orders that arrived with a relative price  $\Delta_x > -s(t)$  was found to decrease as the total depth available on the bid (respectively, ask) side of the LOB increased. A more recent study of the Australian Stock Exchange also reported that the proportion of arriving orders with a relative price of  $\Delta_x \leq -s(t)$  (i.e., market orders) was found to be higher when  $|n(b(t), t)|$  and  $n(a(t), t)$  were larger. By contrast, on the Paris Bourse  $|n(b(t), t)|$  (respectively,  $n(a(t), t)$ ) was found to have little impact on the rate of incoming sell (respectively, buy) orders with a relative price  $\Delta_x \leq -s(t)$ .

By contrast, the distribution of relative prices has been found to be independent of the spread on the LSE (Mike and Farmer, 2008) and the Shenzhen Stock Exchange (Gu *et al.*, 2008b). The distribution of relative prices was also found to be independent of volatility on the Shenzhen Stock Exchange.

On the LSE (Maskawa, 2007), market participants were found to favour placing their limit orders at relative prices similar to those where there was already a large proportion of active orders.

### 3. Arrival rates

On the Stockholm Stock Exchange (Sandås, 2001), order flows at time  $t$  were found to be conditional on both  $L(t)$  and on previous order flows. For the Deutsche Mark/US dollar and Canadian dollar/US dollar currency pairs, traded on foreign exchange markets (Lo and Sapp, 2010), order flows at time  $t$  were found to be conditional on several LOB variables, including  $s(t)$ ,  $n(b(t), t)$  and  $n(a(t), t)$ , depth available behind the best prices, time of day, and recent order flows. However, the precise structure of such conditional dependences was found to vary between the two different currency pairs.

Using several different financial instruments traded in electronic LOBs, Toke (2011) found that arrival rates of limit orders increased on both sides of the LOB following the arrival of a market order. No evidence that market order arrival rates increased following the arrival of a limit order was found.

Using data from 40 stocks on the Paris Bourse, Biais *et al.* (1995) studied the frequencies with which market events belonged to each of 15 different *action classes* (such as “arrival of buy market order”, “arrival of buy limit order within the spread”, and “cancellation of existing sell limit order”), conditional on the action class of the most recently recorded event. The conditional frequency with which a market event belonged to a specified action class, given that the previous market event also belonged to the same action class, was found to be higher than the unconditional frequency with which mar-

ket events belonged to that action class, for all action classes. Such behaviour is known as *event clustering*. The authors offered numerous possible explanations for this phenomenon: market participants may have been strategically splitting large orders into smaller chunks to avoid revealing their full trading intentions (or to minimize market impact, as discussed in Section IV.G); different traders may have been mimicking each other; different traders may have been independently reacting to new information; or different traders may have been trying to *undercut* each other (i.e., cancelling active buy (respectively, sell) orders and resubmitting them at a slightly higher (respectively, lower) price solely to gain price priority). The authors noted that small, successive changes in  $b(t)$  and  $a(t)$  were observed more frequently when  $s(t)$  was large, which they argued provided evidence of undercutting. However, Bouchaud *et al.* (2009) concluded that the phenomenon was driven primarily by strategic order splitting; and found no evidence that different traders mimicked each other’s actions.

On the Swiss Stock Exchange, order flow was found to depend on a number of factors, including volatility, recent order flow, and the state of the limit order book  $L(t)$  (Ranaldo, 2004). Market participants were found to submit more limit orders and fewer market orders during periods when volatility or  $s(t)$  were high. The proportion of orders that arrived with negative relative price was found to decrease as the inter-arrival time between recent orders increased. Market participants were found to submit higher-priced buy orders (respectively, lower-priced sell orders) when there was a greater total depth available on the buy side (respectively, sell side) of the LOB. Buy order submission was found to depend also on the state of the opposite (i.e., sell) side of the LOB, whereas sell order submission was found to depend only on the state of the same (i.e., sell) side of the LOB. Ranaldo noted that such asymmetry may have been caused by market performance over the sample period (the percentage change in  $m(t)$  was positive for all but one of the stocks studied, and exceeded 10% for 4 of the stocks studied), but may also have been caused by buyers and sellers behaving differently from each other in the market.

On the NYSE (Ellul *et al.*, 2003), periods of time with above-average order flow rates were found to cluster together, as were periods with below-average order flow rates. The rate of limit order arrivals was also observed to be higher late in the trading day. The rate of buy (respectively, sell) limit order arrivals was found to increase after periods of positive (respectively, negative) mid-price returns. A similar event clustering as was observed by Biais *et al.* (1995) on the Paris Bourse was also reported. However, the number of occurrences of market events from a specific action class in a given five minute window (and hence the average rate of such occurrences over that window) and the number of occurrences of the market events from the same action class in the previous

five minute window were found to be negatively correlated. Furthermore, the arrival rate of market events from a given action class was found to be more heavily conditional on the action class of the single most recent market event than it was on  $L(t)$ , whereas the distribution of the number of occurrences of market events from a given action class in a given five minute window was found to be more heavily conditional on  $L(t)$  during the previous five minute window than it was on the number of occurrences of market events from any specific action class in the previous five minute window.

On the Australian Stock Exchange (Hall and Hautsch, 2006), the arrival rates of all market events were reported to increase and decrease together. The authors suggested that there might, therefore, have existed other exogenous factors (that they had not measured) that influenced LOB activity overall. In a more recent study of the Australian Stock Exchange (Cao *et al.*, 2008), the arrival rates of market events at time  $t$  was found to be conditional on  $L(t)$ , but not significantly conditional on the state of the LOB at earlier times. This suggests that market participants evaluated only the most recent state of the LOB, and not a longer history, when making order placement and cancellation decisions. No evidence that mid-price returns had a significant impact on order arrival or cancellation rates was found.

#### 4. Cancellations

On the Paris Bourse (Biais *et al.*, 1995), cancellations on each side of the LOB were found to occur more frequently after a matching on that side of the LOB. The authors conjectured that this was evidence of market participants submitting large orders in the hope of finding hidden liquidity, then cancelling any unmatched portions of such orders.

On the Australian Stock Exchange (Cao *et al.*, 2008) priority considerations have been observed to play a key role for market participants when deciding whether or not to cancel their existing active orders. The cancellation rate for active orders was found to increase when new, higher-priority limit orders arrived on the same side of the LOB. In addition, the cancellation rate of active buy (respectively, sell) orders at prices  $p < b(t)$  (respectively,  $p > a(t)$ ) was found to increase when  $n(p - \delta p, t)$  (respectively,  $n(p + \delta p, t)$ ) became zero. The authors suggested that this was because market participants with limit orders at price  $p$  could cancel then resubmit them, at price  $p - \delta p$  (respectively,  $p + \delta p$ ), to possibly gain a better price for the trade (if the limit order eventually matched) without substantial loss of priority. No similar increase in cancellations was observed when  $n(p + \delta p, t)$  (respectively,  $n(p - \delta p, t)$ ) became zero.

#### 5. Price movements

On the Paris Bourse,  $a(t)$  was found to be more likely to decrease (respectively,  $b(t)$  was found to be more likely to increase) immediately after the arrival of a market order that had caused  $b(t)$  to decrease (respectively,  $a(t)$  to increase) (Biais *et al.*, 1995). The authors suggested that such behaviour could be due to market participants reacting to some kind of “information”, either via external sources of news causing a revaluation of the underlying asset or via the downward movement of  $b(t)$  (respectively, upward movement of  $a(t)$ ) itself being interpreted by other market participants as news. Indeed, Potters and Bouchaud (2003) found evidence that on NASDAQ, each new trade (by its very existence) was interpreted by market participants as a piece of new information that had a direct effect on the flow of incoming orders (and, therefore, on prices).

#### 6. Incoming order prices

On the LSE, the relative prices of incoming limit orders were found to be conditional on the bid-price realized volatility per trade (Zovko and Farmer, 2002). More precisely, two time series were constructed by calculating the mean relative price of arriving buy limit orders and the bid-price realized volatility per trade over 10 minute windows. The cross-correlation of the two time series was calculated, and the hypothesis that they were uncorrelated was rejected at the 2.5% level. Changes in bid-price realized volatility were found to immediately precede changes in mean relative price for buy limit orders.<sup>16</sup> A similar behaviour was observed when comparing the time series of ask-price realized volatility and the time series of mean relative price for sell limit orders.

In foreign exchange markets (Lo and Sapp, 2010), during 30 minute windows with high mid-price realized volatility, market participants were found to submit orders with higher relative prices on average.

---

<sup>16</sup> The authors noted that it was not clear from the cross-correlation function alone whether a change in bid-price realized volatility directly caused a change in how market participants chose the relative prices for their buy limit orders shortly thereafter, or whether there was some other external factor that first affected bid-price realized volatility and then affected relative limit prices for buy limit orders. If the latter could be demonstrated, it would support the widely-held belief that many market participants consider realized volatility to be an important factor in making the decision of when to place a limit order (Zovko and Farmer, 2002).

## 7. Order flow

For Canadian stocks (Hollifield *et al.*, 2006), a range of different volatility measures were found to be correlated with order flow rates. For the Deutsche Mark/US dollar and Canadian dollar/US dollar currency pairs, traded on foreign exchange markets, realized mid-price volatility was found to affect order flows (Lo and Sapp, 2010). For US equities traded on Island ECN, and using a variety of different volatility measures, periods of higher volatility were found also to have a lower proportion of limit orders in the arriving order flow (Hasbrouck and Saar, 2002). During such periods, submitted limit orders were found to have an increased probability of execution, and a shorter expected time until execution. Furthermore, on Euronext (Chakraborti *et al.*, 2011b) and for German Index Futures (Kempf and Korn, 1999), mid-price realized volatility was found to increase with the number of arriving market orders in a given time interval. A similar finding was reported in a study of the NYSE (Jones *et al.*, 1994); however, a later study of the NYSE (Elul *et al.*, 2003) reported a positive correlation between higher mid-price realized volatility and the percentage of arriving orders that were limit orders.

On the Australian Stock Exchange, the number of arrivals and cancellations of large limit orders (i.e. whose size was in the upper quartile of the unconditional empirical distribution of order sizes) in a given 5 minute window were found to be positively correlated with mid-price realized volatility during that window and also during the previous 5 minute window (Hall and Hautsch, 2006). However, a more recent study (Cao *et al.*, 2008) found that mid-price realized volatility per trade had only a minimal effect on order flows.

## 8. Limit order book state

A positive (but weak) correlation has been observed between  $s(t)$  and realized mid-price volatility in a wide range of markets (see Wyart *et al.* (2008) and references therein). However, a much stronger positive correlation has been observed between  $s(t)$  and mid-price volatility at the trade-by-trade timescale on the Paris Bourse (Bouchaud *et al.*, 2004), the FTSE 100 (Zumbach, 2004), and the NYSE (Wyart *et al.*, 2008). A recent study of stocks traded on the NYSE (Hendershott *et al.*, 2011) found the once-daily time series of bid-price realized volatility to be positively correlated with the daily mean spread. Stocks with a lower mid price were found to have higher bid-price realized volatility on average. In foreign exchange markets (Lo and Sapp, 2010), the variance of the depth available at any given price was found to increase during periods of high mid-price realized volatility. On the Island ECN (Hasbrouck and Saar, 2002), links relating volatility to various aspects of the depth profile

were investigated, but were all found to be weak.

As discussed in Section III.A, mid-price intra-day volatility on the Sydney Futures Exchange (Bortoli *et al.*, 2006) was found to vary according to how much information about the depth profile market participants had access to in real-time.

## G. Market impact and price impact

A key consideration for a market participant who wishes to buy or sell a large quantity of an asset is how their own actions might affect the LOB (Almgren and Chriss, 2001; Bouchaud *et al.*, 2009; Cont *et al.*, 2011; Eisler *et al.*, 2010; Obizhaeva and Wang, 2005). For example, imagine market participant *A* wishes to buy  $20\sigma$  shares of Company X in the LOB displayed in Figure 6. Submitting a single market order of size  $\omega_x = -20\sigma$  would result in market participant *A* purchasing  $2\sigma$  shares at the price \$1.5438,  $5\sigma$  shares at the price \$1.5439,  $6\sigma$  shares at the price \$1.5440, and  $7\sigma$  shares at the price \$1.5441. However, if market participant *A* were initially to submit only a market order of size  $\omega_x = -2\sigma$ , it is possible that other market participants would submit new limit orders, because by purchasing the  $2\sigma$  shares with highest priority in the LOB, market participant *A* has made it more attractive for other participants to submit new sell limit orders than it was immediately before such a purchase. Market participant *A* could then submit a market order to match to these newly submitted limit orders, and repeat this process until all  $20\sigma$  of the desired shares are purchased. Of course, there is no guarantee that the initial market order of size  $2\sigma$  would stimulate such submissions of limit orders from other market participants. Indeed, it could even cause some other market participants to cancel their existing limit orders or to submit buy market orders, pushing  $a(t)$  up further and thereby ultimately causing market participant *A* to pay a higher price for the overall purchase of the  $20\sigma$  shares. However, such *order splitting* has been empirically observed to be very common in a wide range of different markets (Bouchaud *et al.*, 2009).

The change in  $b(t)$  and  $a(t)$  caused by a market participant's actions is called the *price impact* of the actions. The necessity for market participants to monitor and control price impact predates the widespread uptake of limit order trading. In a quote-driven system, for example, any single market maker only has access to a finite inventory, so there are limits on the size that is available for trade at the quoted prices. Once a market participant reaches these limits, any further trading will have to take place with different market makers, at worse prices. Furthermore, purchasing or selling large quantities of the asset could cause market makers to adjust their quoted prices. Both of these outcomes are examples of price impact. In a limit order market, however, it is possible to con-

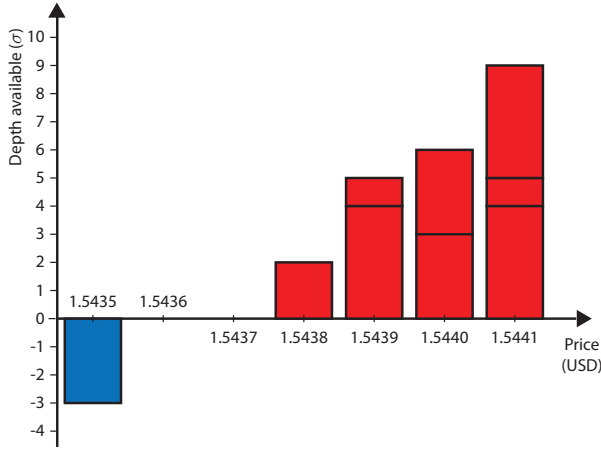


FIG. 6 A LOB for shares in Company X

sider a more general form of impact: the change in the entire LOB state  $L(t)$  that is caused by a market participants' actions. Such impact is called *market impact*. Although, to date, the terms “price impact” and “market impact” have often been used interchangeably to refer only to changes in  $b(t)$  or  $a(t)$ , recent work (Hautsch and Huang, 2009) has shed light on how market participants' actions can affect the depths available at other prices, suggesting that it is appropriate to separate the two notions. Bouchaud *et al.* (2009) provides a detailed review of studies of both price and market impact.

Both price impact and market impact are difficult to quantify formally, as they each consist of two components:

- *Instantaneous (or immediate) impact*: the immediate effects of the specified action.
- *Permanent impact*: the long-term impact due to the specified action causing other market participants to behave differently in the future.

For example, the instantaneous price impact of a market buy order of size  $2\sigma$  in the LOB displayed in Figure 6 is a change in  $a(t)$  from \$1.5438 to \$1.5439. An example of permanent market impact of this buy market order might be another market participant deciding to submit a new sell limit order at a price of \$1.5442.

Instantaneous impact exists because the arrival or cancellation of any order affects  $L(t)$  directly. Bouchaud *et al.* (2009) described three reasons why permanent market impact might exist. First, trades themselves might convey information to other market participants.<sup>17</sup>

Second, agents might successfully forecast short-term price movements and choose their actions accordingly.<sup>18</sup> Third, purely random fluctuations in supply and demand might lead to permanent impact.

The various forms of instantaneous price impact and instantaneous market impact are defined as follows:

**Definition.** The instantaneous bid-price impact of a market event at time  $t'$  is

$$\lim_{t \downarrow t'} b(t) - \lim_{t \uparrow t'} b(t).$$

*Instantaneous ask-price impact and instantaneous mid-price impact are defined similarly.*

**Definition.** The instantaneous bid-price logarithmic return impact of a market event at time  $t'$  is

$$\lim_{t \downarrow t'} (\log b(t)) - \lim_{t \uparrow t'} (\log b(t)).$$

*Instantaneous ask-price logarithmic return impact and instantaneous mid-price logarithmic return impact are defined similarly.*

**Definition.** The instantaneous bid-price impact function  $\phi_b(\omega_x)$  outputs the mean instantaneous bid-price impact for a buy market order of size  $\omega_x < 0$ . The instantaneous ask-price impact function  $\phi_a(\omega_x)$  outputs the mean instantaneous ask-price impact for a sell market order of size  $\omega_x > 0$ . The instantaneous mid-price impact function  $\phi_m(|\omega_x|)$  outputs the mean instantaneous mid-price impact for a market order of size  $|\omega_x|$ .

**Definition.** The instantaneous bid-price logarithmic return impact function  $\Phi_b(\omega_x)$  outputs the mean instantaneous bid-price logarithmic return impact for a buy market order of size  $\omega_x < 0$ . The instantaneous ask-price logarithmic return impact function  $\Phi_a(\omega_x)$  outputs the mean instantaneous ask-price logarithmic return impact for a sell market order of size  $\omega_x > 0$ . The instantaneous mid-price logarithmic return impact function  $\Phi_m(|\omega_x|)$  outputs the mean instantaneous mid-price logarithmic return impact for a market order of size  $|\omega_x|$ .

**Definition.** The instantaneous market impact of a market event at time  $t'$  is

$$\lim_{t \downarrow t'} L(t) \setminus \lim_{t \uparrow t'} L(t).$$

<sup>17</sup> Grossman and Stiglitz (1980) introduced this idea for a general market, and it has since been discussed extensively in a LOB context (see, e.g., (Almgren and Chriss, 2001; Bouchaud *et al.*, 2009; Hasbrouck, 1991; Potters and Bouchaud, 2003)).

<sup>18</sup> This explanation suggests that it is not market participants' actions that cause prices to rise or fall. Instead, such movements happen exogenously and market participants align their actions with them to maximize profits. Bouchaud *et al.* (2009) did not find evidence that this was a good reflection of reality.

It is not possible to quantify precisely the permanent price (respectively, market) impact of an action, because doing so would involve making a comparison between the values of  $b(t)$  and  $a(t)$  (respectively,  $L(t)$ ) if the action hadn't happened and the values (respectively, state) if it had. If the action did happen, it is not possible to know what the values (respectively, state) would have been if it hadn't, and vice-versa.

We now examine instantaneous and permanent price impact in more detail.

### 1. Instantaneous price impact

To date, the study of instantaneous price impact for individual market orders has been conducted primarily via the study of instantaneous price impact and instantaneous logarithmic return impact functions. On the NYSE and American Stock Exchange (Hasbrouck, 1991),  $\phi_m$  was found to be a concave function of  $|\omega_x|$ . This means that the instantaneous price impact of a single market order of size  $|\omega_x|$  was, on average, larger than the sum of the instantaneous price impacts of two market orders  $x_1$  and  $x_2$  of sizes  $|\omega_{x_1}|$  and  $|\omega_{x_2}|$ , such that  $|\omega_{x_1}| + |\omega_{x_2}| = |\omega_x|$ .

Lillo *et al.* (2003) studied the stocks of 1000 different companies traded on the NYSE, and sorted them into 20 groups according to their market capitalization (i.e., the total value of all of a given company's shares). Within each group, they then merged their market data and fitted a single curve to  $\Phi_m(|\omega_x|)$ . They found that  $\Phi_m$  could be fitted with the power law  $\Phi_m(|\omega_x|) = |\omega_x|^\alpha$  for all 20 groups. The value of  $\alpha$  was found to vary between the different groups, taking values between approximately 0.2 and 0.5, although no goodness-of-fit statistics were presented with the results and it is not clear how well the fits performed for individual stocks, rather than aggregated groups.

After performing the change of variables

$$\begin{aligned}\omega'_x &:= \frac{\omega_x}{C^\delta} \\ p' &:= pC^\gamma\end{aligned}$$

where  $C$  is the average market capitalization for stocks in the group and  $\delta$  and  $\gamma$  are fitted constants, they found that  $\Phi_{m'}(|\omega'_x|)$  for each of the 20 groups collapsed onto a single “universal” curve.

A similar collapse of  $\Phi_m$  onto a single (power-law) curve  $\Phi_{m'}(|\omega'_x|) = |\omega'_x|^{0.25}$  was observed for 11 stocks traded on the LSE (Farmer *et al.*, 2005), under the change of variables

$$\begin{aligned}\omega'_x &:= \frac{\omega_x \alpha}{\mu} \\ p' &:= \frac{p \lambda}{\mu}\end{aligned}$$

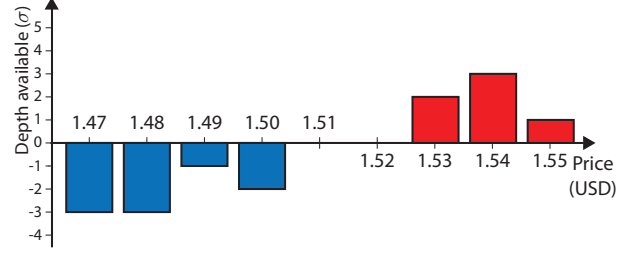


FIG. 7 An example LOB

where  $\mu$ ,  $\lambda$ , and  $\alpha$  were the average arrival rate of market orders, the average arrival rate of limit orders, and the average cancellation rate of active orders per unit size  $\sigma$ , respectively.

Using data from the Shenzhen Stock Exchange, Zhou (2012) partitioned incoming orders according to whether or not they received an immediate full matching. The reason for doing so is as follows. Consider the LOB displayed in Figure 7. If a buy order  $x = (\$1.54, \omega_x)$  arrives, one of following three possible cases will apply:colorblue

1. If  $\omega_x = \sigma$ , then  $n(\$1.54, t) = 2\sigma - \sigma = \sigma$  after its arrival. Therefore,  $b(t)$  and  $a(t)$  remain unchanged, so  $m(t)$  remains unchanged.
2. If  $\omega_x = 2\sigma$ , then  $n(\$1.54, t) = 2\sigma - 2\sigma = 0$  after its arrival. Therefore,  $b(t)$  remains unchanged and  $a(t)$  increases to  $\$1.55$ , so  $m(t)$  increases by 0.005.
3. If  $\omega_x = 2\sigma$ , then  $n(\$1.54, t) = 2\sigma - \omega_x < 0$  after its arrival. Therefore,  $b(t)$  increases to  $\$1.54$  and  $a(t)$  increases to  $\$1.55$ , so  $m(t)$  increases by  $\$0.025$ .

In summary, incoming orders that are fully matched upon arrival always have a strictly smaller instantaneous mid-price impact than those orders that are not. This was reinforced by the data from the Shenzhen Stock Exchange, where incoming orders that were only partially matched upon arrival were found to have much larger instantaneous mid-price impact than incoming orders that were fully matched upon arrival. Furthermore,  $\Phi_m(\omega_x)$  was found to take a different functional form in the two cases:

- For incoming orders that were only partially matched upon arrival,  $\Phi_m(|\omega_x|)$  was found to be constant for all  $|\omega_x| < 10000$  shares, then to increase for larger values of  $|\omega_x|$ .
- For incoming orders that were fully matched upon arrival,  $\Phi_m(|\omega_x|)$  was reported to follow the power law  $\Phi_m(|\omega_x|) = A|\omega_x|^\alpha$ , where  $A$  is a constant that varied from stock to stock. Among buy orders, the mean value of  $\alpha$  was  $0.66 \pm 0.05$ . Among sell orders, the mean value of  $\alpha$  was  $0.69 \pm 0.06$ .



After applying the change of variables

$$\Phi'_m(|\omega_x|) := \frac{\Phi_m(|\omega_x|)}{\langle \Phi_m \rangle}$$

$$\omega'_x := \frac{\omega_x}{\langle |\omega_x| \rangle}$$

where the angle brackets  $\langle \cdot \rangle$  denote the mean value taken across all incoming market orders in the data, Zhou found that the  $\Phi'_m(\omega'_x)$  curves for all studied stocks collapsed onto a single curve for incoming orders that were fully matched upon arrival, and onto another single curve for incoming orders that were only partially matched upon arrival. In particular, the asymmetry between the bid side and the ask side of the LOB was no longer present after the rescaling.

On both the Paris Bourse and NASDAQ (Potters and Bouchaud, 2003), a logarithmic functional form was found to provide a better fit to  $\phi_m$  than was a power-law relationship. Furthermore, power-law relationships were found to overestimate the mean instantaneous mid-price impact of very large market orders on both the LSE and the NYSE (Farmer and Lillo, 2004).

## 2. Permanent price impact

As discussed above, it is impossible to exactly quantify the permanent price impact of a market event. However, to gain some insight into the longer-term effects of market events, several empirical studies have compared changes in  $b(t)$  and  $a(t)$  over specified time intervals with measures of *trade imbalance*.

**Definition.** The trade imbalance count during time interval  $T = [t_1, t_2]$ , denoted  $\Omega_c(T)$ , is the difference between the total number of incoming buy market orders and the total number of incoming sell market orders that arrive during time interval  $T$ .

**Definition.** The trade imbalance size during time interval  $T = [t_1, t_2]$ , denoted  $\Omega_\omega(T)$ , is the difference between the total absolute size of all incoming buy market orders and the total size of all incoming sell market orders that arrive during time interval  $T$ .

Using data for Deutsche Mark/US dollar and US dollar/Yen foreign exchange markets, Evans and Lyons (2002) performed an ordinary least squares regression, comparing the ask-price logarithmic return between successive trading days and the daily trade imbalance count. A statistically significant, positive, linear relationship was found between the two variables.

For German Stock Index futures, the average mid-price logarithmic return over a 5 minute window was found to be a concave function of the trade imbalance count during that window (Kempf and Korn, 1999). For the largest

100 stocks on the NYSE (Gabaix *et al.*, 2006), the average mid-price logarithmic return was found to follow the relationship  $\Omega_\omega(T)^{0.5}$  for time intervals of length 15 minutes. For the 116 most liquid stocks in the US between 1994–1995 (Plerou *et al.*, 2002), for a variety of different time interval lengths, the average change in mid-price over the interval was found to be a concave function of  $\Omega_\omega(T)$ . Furthermore, for small values of  $\Omega_\omega(T)$ , the average change in mid-price over the interval was found to be well approximated by  $\Omega_\omega(T)^\alpha$ , with the value of  $\alpha$  depending on the length of the time interval, and ranging from  $\alpha = 1/3$  for intervals of length 5 minutes to  $\alpha = 1$  for intervals of length 195 minutes. Similarly, for the Astrazeneca stock (traded on the LSE) (Bouchaud *et al.*, 2009), the average mid-price logarithmic return was found to become an increasingly linear function of the length of time interval  $T$  as the length of  $T$  was increased.

Cont *et al.* (2011) recently proposed that in limit order markets, price impact should be studied as a function of the difference between limit order flow on the bid side and the ask side of the LOB, rather than of  $\Omega_\omega(T)$ , thus acknowledging that cancellations can also have price impact. Using data from the NYSE, the authors performed an ordinary least squares regression of the mean change in mid-price over a time window of length 10 seconds onto the limit order flow imbalance over the same time window, for each stock separately. For 43 out of the 50 stocks studied, the gradient coefficient of the regression was found to be significantly different from 0, working at the 95% level. The value of the gradient coefficient was found to be larger on average for those stocks with smaller mean values of  $|n(b(t), t)|$  and  $n(a(t), t)$ . The authors noted that their ordinary least squares regressions provided a “surprisingly high” strength of fit across all stocks, despite the nuances of how the individual stocks were traded. In particular, the mean of the  $R^2$  coefficients for the individual stocks’ regressions was found to be 65%. When the regressions were repeated, but using  $\Omega_\omega(T)$  rather than limit order flow imbalance as the independent variable, the average of the  $R^2$  coefficients was only 32%. Furthermore, the authors conjectured that any observable relationship between price impact and  $\Omega_\omega(T)$  was actually a by-product of the fact that  $\Omega_\omega(T)$  was correlated with limit order flow imbalance, and that limit order flow imbalance was the real driving force.

## 3. Market impact

In contrast to the wealth of empirical studies on price impact, little has been published to date on market impact. However, by considering the impact of market events on the depths available not only at  $b(t)$  and  $a(t)$  but also at nearby prices, Hautsch and Huang (2009) were able to track how order arrivals affected the state of

the LOB  $L(t)$  of 30 stocks traded on Euronext. In particular, limit orders placed at a price  $p \leq b(t)$  (respectively,  $p \geq a(t)$ ) were found to cause a significant permanent increase in  $b(t)$  (respectively, decrease in  $a(t)$ ), working at the 95% level. Limit orders with a relative price of  $\delta p$  or  $2\delta p$  were found to affect  $b(t)$  and  $a(t)$  only 20% less, on average, than limit orders placed at  $b(t)$  and  $a(t)$  (i.e., with a relative price of 0). Limit orders placed with negative relative price were found to have significant market impact. The market impact of a market order was found to be, on average, four times greater than that of a limit order of the same size.

The effect on  $b(t)$  and  $a(t)$  caused by limit orders arriving with a relative price of zero was observed to occur more quickly than for limit orders arriving with positive relative price. The market impact of limit orders placed inside the spread was found to be largely instantaneous,<sup>19</sup> with little permanent impact, whereas limit orders that arrived with zero or positive relative price were observed to have no immediate impact but significant permanent impact.

Similar results were reported for all stocks studied, but asymmetries were found between the bid side and the ask side of the LOB (much like Kempf and Korn (1999) found for price impact). The authors conjectured that the impact they observed was due partly to arriving orders triggering an instantaneous imbalance in supply and demand, and partly to other market participants interpreting order arrivals as containing information, thereby causing them to adjust their own future actions and leading to permanent market impact. The results suggested that the arrivals of market orders were interpreted by market participants as being a particularly strong information signal. Such an observation provides a possible explanation as to why so many market participants choose to place iceberg orders: placing an iceberg order is an effective way to hide the true size of limit orders from the market, and thus to minimize market impact.

## H. Stylized facts

Several non-trivial statistical regularities have been detected in empirical data from a wide range of different markets. Such regularities have come to be known as the *stylized facts* of markets, and they may provide interesting insights into the behaviour of market participants (Cont, 2001) and the structure of markets themselves (Bouchaud *et al.*, 2009). The stylized facts are also useful from a modelling perspective, because a model's inability to reproduce one or more stylized facts can be

used as an indicator for how it needs to be improved, or used as justification for ruling it out altogether. For example, the existence of volatility clustering eliminates the simple random walk as a model for the evolution of  $m(t)$  through time, as the existence of volatility clustering in real mid-price time series implies that large price variations are more likely to follow large price variations than they are to occur unconditionally (Lo and MacKinlay, 2001).

Reproduction of the stylized facts has proven to be a serious challenge for LOB models (Chakraborti *et al.*, 2011b), particularly for those based on zero-intelligence assumptions, which have, thus far, produced more volatile price series than empirical observations have suggested is appropriate (Chakraborti *et al.*, 2011a). This might imply that the strategic behaviour of real market participants somehow stabilizes prices, and is therefore an important ingredient in real LOB trading.

Cont (2001) reviewed a wide range of stylized facts and their estimation from empirical data; here we discuss a small subset that we consider to be the most relevant from a LOB perspective. The stylized facts presented here are of particular theoretical interest because they all suggest that non-equilibrium behaviour plays an important role in LOBs. A result from statistical mechanics is that systems that are in equilibrium give rise to distributions from the exponential family (Mike and Farmer, 2008), whereas the distributions describing several aspects of LOB behaviour exhibit power-law tails, highlighting the possibility that LOBs may always be in a transient state.

### 1. Heavy-tailed return distribution

Over all timescales ranging from seconds to days, the unconditional distribution of mid-price returns displays tails that are heavier than a normal distribution (i.e., they have positive excess kurtosis). The exact form of the distribution has been found to vary with the timescale used. Across a wide range of different markets (e.g., (Gopikrishnan *et al.*, 1998; Gu *et al.*, 2008a)), at the shortest timescales the tails of the distribution appear to be well-approximated by a power law with exponent  $\alpha \approx 3$ , thus earning the name “the cubic law of returns” in the literature. Stanley *et al.* (2008) went on to conjecture that such a universal power-law tail might be a consequence of power-law tails in both the distribution of market order sizes and the instantaneous mid-price logarithmic return impact function. However, Mu and Zhou (2010) reported that this relationship did not hold in emerging markets, and Drożdż *et al.* (2007) noted that the tails are actually thinner, i.e.,  $\alpha > 3$ , in the most recent data, highlighting that the quantitative form of stylized facts may themselves change over time as trading styles evolve. At longer timescales, the distribution

<sup>19</sup> Note that, by definition, a buy (respectively, sell) limit order placed inside the spread immediately affects  $b(t)$  (respectively,  $a(t)$ ).

becomes increasingly better approximated by a normal distribution (a behaviour often referred to as *aggregational Gaussianity*) (Cont, 2001; Gopikrishnan *et al.*, 1999; Zhao, 2010).

The return distribution has been found to exhibit heavy tails on Euronext (Chakraborti *et al.*, 2011b), the Paris Bourse (Plerou and Stanley, 2008), the S&P 500 index (Cont, 2001; Gallant *et al.*, 1992; Gopikrishnan *et al.*, 1999), foreign exchange markets (Guillaume *et al.*, 1997), the NYSE (Gopikrishnan *et al.*, 1998), the American Stock Exchange (Gopikrishnan *et al.*, 1998; Plerou and Stanley, 2008), NASDAQ (Gopikrishnan *et al.*, 1998), the LSE (Plerou and Stanley, 2008), and the Shenzhen Stock Exchange (Gu *et al.*, 2008a). The heavy-tailed nature of returns has particularly important consequences for market participants, as it highlights that large movements in price are more likely than they would be if returns were normally distributed, and is central to the management of risk when calculating investment strategies.

## 2. Autocorrelation of returns

Except on very short timescales, when it exhibits weak negative autocorrelation, the time series of mid-price returns does not display any significant autocorrelation (Chakraborti *et al.*, 2011b; Cont, 2005; Stanley *et al.*, 2008). As highlighted by Cont (2001), the absence of autocorrelation in returns can be explained using perfect-rationality arguments. If returns were indeed autocorrelated, rational market participants would employ simple strategies that used this fact to generate positive expected earnings. Such actions would themselves reduce the level of autocorrelation, so autocorrelation would not persist.

The absence of autocorrelation on all but the shortest timescales is a well-established empirical fact that has been observed in a very large number of markets, including the NYSE (Aït-Sahalia *et al.*, 2005; Cont, 2005), Euronext (Chakraborti *et al.*, 2011b), the US dollar/Yen and Pounds Sterling/US dollar currency pairs traded on foreign exchange markets (Bouchaud and Potters, 2003; Cont *et al.*, 1997), the S&P 500 (Bouchaud and Potters, 2003; Gopikrishnan *et al.*, 1999), German interest rates futures contracts (Bouchaud and Potters, 2003), and crude oil futures (Zhao, 2010).

The timescales over which the negative autocorrelation persists is less clear, however. Using data from 1984–1996, the S&P500 exhibited negative autocorrelation in mid-price returns on timescales of up to around 20 minutes (Gopikrishnan *et al.*, 1999). However, using data from 1991–2001, the S&P 500 showed negative correlation up to 10 minutes, but not 15 minutes. Over the same date range, German interest rates futures contracts showed similar behaviour, whereas the Pounds Sterling/US dollar currency pair exhibited negative au-

tocorrelation up to timescales of 20 minutes, but not 25 minutes (Bouchaud and Potters, 2003). On the NYSE, a publication from 2005 reported that negative autocorrelation persisted at 5 minutes, but not 10 minutes (Cont, 2005) (although no exact date of when the data itself was collected was given). However, data from Euronext during 2007–2008 exhibited no significant autocorrelation over time windows of 1 minute (Chakraborti *et al.*, 2011b). Furthermore, in NYSE data from 2010, no significant autocorrelation was present over any timescales of 20 seconds or longer (Cont *et al.*, 2011), and for crude oil futures contracts traded in 2005 (Zhao, 2010), negative autocorrelation was found to persist for only 10–15 seconds. In summary, it appears that any negative autocorrelation disappears more quickly in more recent market data than it does in older data, again indicating that the exact quantitative details of stylized facts may have changed over time.

## 3. Long memory

Several empirical studies have reported LOB time series to exhibit long memory (as defined in Section III.M).

Time series of absolute or square mid-price returns display positive long memory over timescales as long as weeks or even months (Cont, 2001; Liu *et al.*, 1997; Stanley *et al.*, 2008). More precisely, the square mid-price returns for S&P 500 index futures (Cont, 2001), the NYSE (Cont, 2005), the US dollar/Yen currency pair (Cont *et al.*, 1997) and crude oil futures (Zhao, 2010) were all found to exhibit slowly-decaying positive autocorrelations at intra-day timescales, as were the absolute mid-price returns on the Paris Bourse (Chakraborti *et al.*, 2011b) and the Shenzhen Stock Exchange (Gu and Zhou, 2009a). The Hurst exponent  $H$  of the volatility series was reported to be  $H \approx 0.8$  on the Paris Bourse,  $H \approx 0.815$  for the US dollar/Yen currency pair, and  $H \approx 0.58$  on the Shenzhen Stock Exchange.

The time series constructed by assigning the value +1 to incoming buy orders and −1 to incoming sell orders has been found to exhibit a long memory on the Paris Bourse (Bouchaud *et al.*, 2004), the NYSE (Lillo and Farmer, 2004), and the Shenzhen Stock Exchange (Gu and Zhou, 2009a). On the LSE (Bouchaud *et al.*, 2009; Lillo and Farmer, 2004; Mike and Farmer, 2008), the time series constructed by assigning the value +1 to incoming buy market orders and −1 to incoming sell market orders has been found to exhibit a long memory, as has the same series for buy or sell limit orders and for buy or sell active order cancellations. Statistically significant differences were reported between the values of the exponents for different stocks. Volatility,  $n(b(t), t)$ , and  $n(a(t), t)$  were all also found to exhibit long memory. The arrival of external news and the strategic splitting of orders (see Section IV.G) were both offered as potential causes of such long

memory. Zovko and Farmer (2002) also reported that on the LSE, the time series of relative prices of submitted limit orders was found to be a long-memory process, with Hurst exponent  $H = 0.8$ . A similar long-memory effect was reported on the Shenzhen Stock Exchange (Gu and Zhou, 2009a), with Hurst exponent  $H = 0.62$ .

As Lillo and Farmer (2004) discussed, it might be expected that the various forms of long memory observed would lead to high levels of predictability in mid-price returns. However, anti-correlations were found to exist between the different long memory processes, thereby removing exploitable predictability in the mid-price return series. In particular, the long memory of market order arrivals was found to be offset by the long memory in  $n(b(t), t)$  and  $n(a(t), t)$ , meaning that when predictability of market order arrivals was high, the probability that a buy (respectively, sell) market order would cause a change in  $m(t)$  was low, because  $|n(b(t), t)|$  and  $n(a(t), t)$  were larger. Bouchaud *et al.* (2004) offered an alternative explanation for the absence of such predictability, instead suggesting that the long memory in the arrival of limit orders offset the long memory in the arrival of market orders, thereby again making large changes in  $m(t)$  unlikely.

## V. LIMIT ORDER BOOK MODELS

During recent years, the economics and physics communities have both made substantial progress with LOB modelling, as surveyed in (Chakraborti *et al.*, 2011a; Parlour and Seppi, 2008). However, work by the two communities has remained largely independent (Farmer *et al.*, 2005). Work by economists has tended to be “trader-centric”, using perfect-rationality frameworks to derive optimal trading strategies given certain market conditions. These models have generally treated order flow as static. By contrast, models from physics have tended to be conceptual toy models of the evolution of  $L(t)$ . By relating changes in order flow to properties of the LOB, these models treat order flow as dynamic (Farmer *et al.*, 2005). Both approaches clearly have their strengths. An understanding of trading strategies is crucial for market participants and regulators of markets alike (Alfonso *et al.*, 2010; Almgren and Chriss, 2001; Cao *et al.*, 2008; Evans and Lyons, 2002; Foucault *et al.*, 2005; Gatheral, 2010; Goettler *et al.*, 2006; Hall and Hautsch, 2006; Holfield *et al.*, 2006; Roşu, 2010; Sandås, 2001; Seppi, 1997; Wyart *et al.*, 2008). An understanding of the state of the LOB and order flow helps to explain many of the regularities found in empirical data, and gives some insight into whether such regularities might emerge as a consequence of market microstructure, rather than the strategic behaviour of those trading within it (Bouchaud *et al.*, 2009; Farmer *et al.*, 2005; Gu and Zhou, 2009a; Mike and Farmer, 2008; Smith *et al.*, 2003).

In this section, we assess existing LOB models in terms of their ability to accurately mimic the structure of limit order trading and reproduce empirical facts from LOB data (as discussed in Section IV), and we highlight the main challenges and problems that are yet to be addressed.

### A. Perfect-rationality approaches

In the traditional economics approach, rational investors faced with straightforward buy or sell possibilities choose portfolio strategies of holdings to maximize personal utility, subject to budget constraints. However, limit order markets provide a substantially more complicated scenario. Rather than submitting orders for exact quantities at exact prices, an investor calculating the ideal portfolio may attempt to construct this using both limit orders and market orders. Investors can be certain about the state of their portfolio if they only submit market orders, but the inherent uncertainty of execution of limit orders adds substantially to the uncertainty of the evolving state space for all investors who choose to use them. Furthermore, for a market participant to successfully calculate fill probability distributions for limit orders, it is necessary to condition on  $L(t)$ . If a market participant also believes that there are regularities present in order flow, these must also be conditioned on. However, in heavily-populated markets it is unlikely that any such regularities will persist, as they could provide statistical arbitrage opportunities if they did so.

#### 1. Cut-off strategies

Many early perfect-rationality models aimed to address market participants’ decision making via the use of a *cut-off strategy*, analogous to a hypothesis test in statistical inference:

**Definition.** *When attempting to choose between decision  $D_1$  and decision  $D_2$  at time  $t$ , an individual employing a cut-off strategy compares the value of a statistic  $Z(t) \in \mathbb{R}$  with a cut-off point  $z \in \mathbb{R}$ , and makes the decision*

$$\begin{aligned} D_1, & \text{ if } Z \leq z \\ D_2, & \text{ otherwise.} \end{aligned}$$

The statistic  $Z(t)$  can be any statistic related to  $L(t)$ , current or recent order flow, the actions of other market participants, and so on. For example, a market participant who wishes to place a buy order at time  $t$  might decide to submit a buy market order if  $s(t)$  is smaller than  $5\delta p$ , or to submit a buy limit order with a relative price of  $\Delta = -\delta p$  otherwise.

Cut-off strategies are not sufficiently rich to model situations where market participants choose outcomes from

a range of values, rather than a binary set. Nevertheless, cut-off strategies have been used widely in perfect-rationality models, as they drastically reduce the dimensionality of the decision space available to market participants, which is very appealing from the standpoint of tractability.

To our knowledge, the first model that addressed endogenous decision making between limit and market orders in a setting that resembled limit order trading was due to Chakravarty and Holden (1995). Trading was modelled as a single-period game. First, a market maker arrived and set quotes. Then, all other market participants arrived simultaneously and chose between submitting limit or market orders, using a cut-off strategy based on the difference between their private valuation of the asset and the quotes set by the market maker. Finally, all trades were executed simultaneously using pro-rata priority (there is no concept of time priority in a single-period framework). Although this single-step set-up fails to capture many crucial features of real limit order trading and greatly simplifies the decision process facing market participants, it demonstrated that optimal strategies for informed traders could involve submitting either limit orders or market orders, depending on how the market maker acted. This in turn highlighted that endogenous order choice for market participants was a crucial feature of a successful LOB model.

Foucault (1999) extended this idea by modelling trading as a multi-step game in which market participants were assumed to arrive sequentially. Limit orders were assumed to remain in the LOB for only one period; if the next arriving market participant did not submit a market order to match to an existing limit order, it would expire and be removed from the LOB. Upon arrival, each market participant chose between placing a limit order or a market order, then left the market forever. After each such departure of a market participant, the game ended with some fixed probability; otherwise, a new market participant arrived and the process repeated. Foucault showed that in such a game, the optimal trading strategy for a market participant was a cut-off strategy based on the market participant's private valuation of the asset and the price of the existing limit order (if one existed at all). More precisely, market participants who observed that the LOB already contained an active buy (respectively, sell) order at a price above (respectively, below) their cut-off price should submit a market order to perform an immediate trade, otherwise they should submit a limit order instead.

Foucault's model contains several assumptions that poorly mimic important aspects of real limit order trading, including the assumptions that limit orders only last for a single period of time and that there is a random, exogenous stopping time governing trading. These assumptions restrict the model's ability to make realistic predictions about order flow dynamics, and, therefore,

about how market participants estimate order fill probabilities when deciding how to act. However, Foucault's model highlighted how the probability that a submitted limit order becomes matched depended explicitly on future market participants' actions, which are themselves endogenous. Furthermore, it showed that in deriving their own optimal strategies, market participants must actively consider the strategies of the other market participants.

Parlour (1998) extended Foucault's model by removing the assumption that limit orders expired after a single period. Although Parlour's model only allowed limit orders to be submitted at one specific price and did not allow limit order cancellations (thus greatly simplifying the trading process (Hollifield *et al.*, 2006)), the work identified explicit links between market participants' strategies and  $L(t)$ . In particular, Parlour demonstrated that the optimal decision between submitting a limit or a market order should be made by employing a cut-off strategy, in which newly-arriving market participants assessed both sides of the LOB, not just the side that they would be directly affecting, in order to estimate the fill probability for a limit order. If the estimated fill probability was sufficiently high, the market participant should submit a limit order; otherwise submitting a market order. Parlour argued that limit orders became less attractive later in the trading day due to their lower fill probabilities before the day ended.

Hollifield *et al.* (2004) empirically tested whether cut-off strategies for choosing between limit and market orders (such as those discussed above) could explain the observed actions of market participants trading the Ericsson stock on the Stockholm Stock Exchange. Working at the 1% level, the hypothesis was accepted when the bid side or the sell side of the LOB were each considered in isolation but rejected when both sides of the LOB were considered together, due to the existence of several limit orders with extremely low fill probabilities whose expected payoff was too low for the model to justify. Hollifield *et al.* concluded that cancellations, which were absent from the models discussed above, must play an important role in real LOBs. Hollifield *et al.* (2006) later produced a model with endogenous cancellations and concluded that optimal decisions were not made via a simple cut-off strategy, but rather depended on the evaluation of several interacting functions that varied between market participants according to their personal waiting-time penalty function.

## 2. Fundamental values and informed traders

Some perfect-rationality models centre around the idea that a subset of market participants are *informed traders* who know the "fundamental" or "true" value for the asset being traded; everyone else is *uninformed* and does not

know this true value (see e.g., (Copeland and Galai, 1983; Glosten, 1994; Glosten and Milgrom, 1985; Kyle, 1985)). Bouchaud *et al.* (2009) noted that many researchers now reject the idea of assets having fundamental values, but such models do provide insight into price formation in markets with asymmetric information.

In the classic Kyle (1985) model, uninformed traders placed limit and market orders to trade with each other. At the same time, informed traders observed the LOB and, if ever an uninformed trader posted a buy limit order with a price above (respectively, sell limit order with a price below) the fundamental value, an informed trader would submit a market order to “pick-off” the mispriced limit order and thereby make a profit.

However, more recent models have highlighted several reasons that informed traders should sometimes choose to submit limit orders rather than market orders, for example to avoid detection by other market participants (who would surely mimic an informed trader whom they knew to be well-informed about the fundamental value of the asset, and thereby erode his/her profit opportunities (Roşu, 2010)) and to obtain better prices for their trades (Chakravarty and Holden, 1995; Roşu, 2010).

An example of such a model is Goettler *et al.* (2006). In a limit order market populated by agents who act upon asymmetric information, market participants were assumed to arrive following a Poisson process. Upon arrival, a market participant submitted any desired orders, choosing freely among prices. The agent then left the market and re-arrived following an independent Poisson process. Upon rearrival, the agent had the option to cancel or modify any of their active orders. When a market participant performed a trade, they left the market forever. Additionally, any market participant could, at any time, pay a fee to become informed about the fundamental value of the asset, and to stay informed until they eventually traded. Goettler *et al.* investigated when it was optimal for a market participant to purchase the information (if at all), and found that a market participant’s willingness to do so should decrease as their desire to trade increased. They found that speculators, who trade purely for profit, should buy the information the most often, and that the value of the information increased with volatility. The optimal strategy for an informed market participant was found to contain submissions of both limit orders and market orders. However, as Parlour and Seppi (2008) discussed, Goettler *et al.*’s step forward in realism came at the cost of discarding analytical tractability, forcing the authors to rely solely on numerical computations rather than closed-form expressions.

Roşu (2010) also investigated how informed traders should optimally choose between limit orders and market orders. In the absence of cancellation costs and monitoring costs (so that all perfectly-rational market participants continuously monitored and actively updated

all of their existing limit orders), it was shown that if  $b(t)$  or  $a(t)$  exhibited an extreme mispricing, an informed trader should submit a market order, to capitalize on the mispricing before any other informed market participants with the same information completed the trade first. However, if  $b(t)$  or  $a(t)$  exhibited a less extreme mispricing, a limit order should be submitted instead (to gain a better price for the trade, if it was matched). The price impact of a single informed trader’s order submissions was found to be insufficient to reset  $b(t)$  and  $a(t)$  to the fundamental levels as described by the information, so any subsequent informed market participants who arrived at the market with the same information were able to perform similar actions to also make a profit. Roşu argued that this was a possible explanation for the empirically observed phenomenon of event clustering (as discussed in Section IV.F.3).

Roşu (2009) replaced the idea that market participants who selected different prices for their orders must have done so due to asymmetric information (Glosten and Milgrom, 1985; Kyle, 1985) with the notion that different market participants might have selected different prices for their orders because they valued the immediacy of trading differently. For example, in real markets some market participants need to trade immediately and therefore submit a market order; others do not need to trade immediately and can therefore submit a limit order in the hope of eventually trading at a better price. In Roşu’s model, market participants could modify and cancel their active orders in real time, making it the first perfect-rationality LOB model to reflect the full range of actions available to market participants. Rather than complicating the model, Roşu demonstrated that limit order cancellations simplified the decision-making problem. He proved the existence of a unique Markov-perfect equilibrium in the game and derived the optimal strategy for a newly arriving market participant. Furthermore, he showed that in a LOB populated by market participants following such a strategy, a hump-shaped depth profile would emerge, in agreement with empirical findings from a number of different markets (see Section IV.D).

### 3. Minimizing market impact

As discussed in Section IV.G, determining how to minimize the market impact of an order is a key consideration for market participants. Several perfect-rationality models have suggested that the event clustering found in empirical data (as outlined in Section IV.F.3) might be a signature of market participants attempting to minimize their market impact when executing large orders (Bouchaud *et al.*, 2009). Lillo *et al.* (2005) showed that the power-law decaying autocorrelation function exhibited by order flows present in empirical data could be reproduced by a model in which market participants who

wished to buy or sell a large quantity of the asset did so by submitting a collection of smaller orders sequentially over some period of time.

In a discrete time framework, Bertsimas and Lo (1998) derived an optimal trading strategy for a market participant seeking to minimize expected trading costs, including those due to market impact, when processing a very large order that had to be completed in the next  $k$  time steps. They showed that if prices followed an arithmetic random walk, then the original order should be split into  $k$  equal blocks and submitted uniformly through time. Additionally, they showed that if prices reflected some form of exogenous information that was serially correlated through time, the optimal strategy involved dynamically adjusting trade quantities at every step. However, both of these assumptions about the behaviour of prices poorly mimic the structure of empirically observed price series (Lo and MacKinlay, 2001). Almgren and Chriss (2001) derived a similar strategy for executing a large order, but instead by maximizing the utility of trading revenues including a component to penalize for uncertainty.

Obizhaeva and Wang (2005) considered the optimal execution problem in continuous time. In this set-up, choosing optimal times to submit orders, not just their sizes, is crucial. The authors demonstrated that simply considering the limit  $k \rightarrow \infty$  of a  $k$ -period discrete-time model was not appropriate, as it led to a degenerate solution where execution costs were strategy-independent. By making some strong assumptions about the LOB, including assuming that after the arrival of a market order, the depth profile underwent exponential recovery through time<sup>20</sup> back to a neutral uniform state of  $n(p, t) = n(p', t)$  for all prices  $p, p'$ , Obizhaeva and Wang derived explicit optimal execution strategies and concluded that the theoretical optimum required the submission of uncountably many orders over a finite time period. Alfonsi *et al.* (2010) developed the model by removing the assumption that the “neutral” state of the depth profile must be uniform, although recovery to the neutral state was still assumed to be exponential. In the absence of considerations of permanent impact, Alfonsi *et al.* showed that in discrete (respectively, continuous) time, the optimal execution strategy involved initially submitting a large market order to “stimulate recovery”, small equally-sized market orders at each intermediate time step (respectively, at a fixed rate in continuous time), then another large market order at the end. When permanent impact was also considered, the problem was solved for the special case of a uniform neutral depth profile.

<sup>20</sup> Such recovery of the depth profile, often known as its *resiliency*, has been discussed in both the empirical (Biais *et al.*, 1995; Bouchaud *et al.*, 2004; Potters and Bouchaud, 2003) and modelling literature (Foucault *et al.*, 2005; Roşu, 2009).

## B. Zero-intelligence approaches

As noted above, most perfect-rationality models to date have relied on a series of auxiliary assumptions in order to quantify unobservable parameters, making it difficult to relate their predictions to real limit order markets. By contrast, zero-intelligence models work in a purely quantitative framework, the general approach being to construct a stochastic model for observable processes such as order arrivals and cancellations, estimate its parameters from historical data, use it to produce simulated output, and then test whether the output agrees with empirical regularities observed in real LOB data (such as the mean depth profile, the spread distribution, and the stylized facts discussed in Section IV.H). In such a framework, falsifiable hypotheses can be formulated and the predictive power of models can be measured by training them on a subset of available data (“in-sample”). The model’s output may then be evaluated against other data (“out-of-sample”).

### 1. Model framework

Most zero-intelligence LOB models use the framework introduced by Bak *et al.* (1997) to model the evolution of  $L(t)$ . Orders are modelled as particles on a discrete one-dimensional lattice, whose locations correspond to price. Each particle corresponds to an order of size  $\sigma$ , so an order of size  $k\sigma$  is modelled by  $k$  separate particles. Sell orders are represented as a particle of type  $A$  and buy orders are represented as a particle of type  $B$ . When two orders of opposite type occupy the same point on the pricing grid, the annihilation  $A + B \rightarrow \emptyset$  occurs. Figure 8 illustrates how a LOB is modelled in this way.

### 2. Random-walk diffusion models

Bak *et al.* (1997) introduced the earliest class of zero-intelligence LOB models, in which  $L(t)$  was modelled as particles diffusing along the price lattice. Given an initial LOB state (with all  $A$  particles to the right of all  $B$  particles), each particle was assumed to undergo a random walk along the price lattice. If two particles of opposite type occupied the same point on the price lattice, an annihilation occurred. Initially, such models were studied analytically and via Monte Carlo simulation (Bak *et al.*, 1997; Chan *et al.*, 2001; Eliezer and Kogan, 1998; Tang and Tian, 1999), and produced several possible explanations for empirical regularities observed in real LOB data, such as the hump shaped depth profile, as discussed in Section IV.D. However, they have since been rejected by the modelling community because several subsequent studies have concluded that the diffusion of active orders across different prices is not ob-

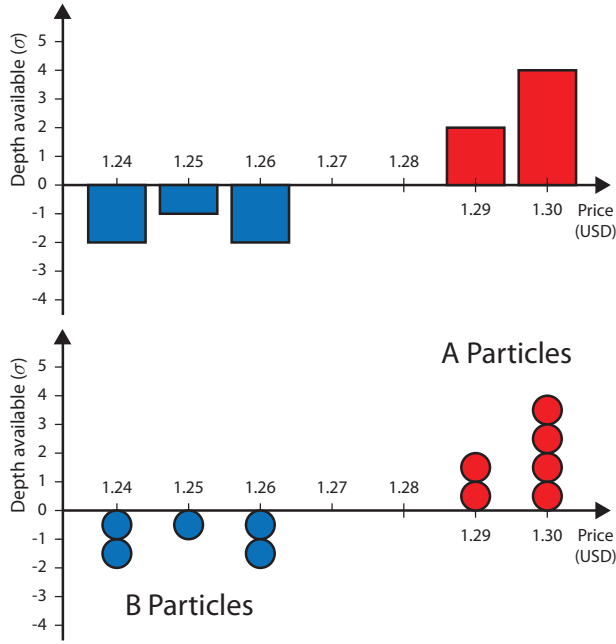


FIG. 8 A LOB and its corresponding representation as a system of particles on a pricing lattice

served in real LOBs (Chakraborti *et al.*, 2011a; Challet and Stinchcombe, 2001; Farmer *et al.*, 2005). Nonetheless, these models sparked the idea that empirical regularities in real LOB data that were previously thought to be a direct consequence of market participants' strategic actions could be reproduced in a zero-intelligence framework. This idea has subsequently become a central theme of the zero-intelligence modelling literature (Bouchaud *et al.*, 2009; Farmer and Foley, 2009; Farmer *et al.*, 2005; Smith *et al.*, 2003).

### 3. Discrete-time models

Maslov (2000) introduced a model that bore a much stronger resemblance to real limit order trading than the price diffusion models discussed above. In Maslov's model, a single market participant was assumed to arrive at the limit order market at each discrete time step. With probability  $\frac{1}{2}$ , this market participant was a buyer; otherwise, they were a seller. Independently of whether they were a buyer or a seller, with probability  $1 - r$  the market participant submitted a market order; otherwise, they submitted a limit order  $x = (p_x, \sigma)$ , with

$$p_x = p' + \pi K$$

where  $p'$  was the most recent price at which a matching had occurred,  $K$  was a random variable with a specified distribution, and

$$\pi = \begin{cases} -1, & \text{if the trader was a buyer,} \\ 1, & \text{if the trader was a seller.} \end{cases}$$

No cancellations or modifications to active orders were allowed. Even with only 1000 iterations and in very simple set-ups (such as  $r = \frac{1}{2}$  and  $K = 1$  with probability 1; or  $r = \frac{1}{2}$  and  $K \sim \text{Uniform}\{1, 2, 3, 4\}$ ), the series of prices at which trades occurred was found to exhibit negative autocorrelation on event-by-event timescales and a heavy-tailed return distribution. Slanina (2001) showed that the negative autocorrelation and heavy-tailed return distribution remained when implementing a mean-field approximation to replace the tracking of prices of individual limit orders with that of a mean value that increased when a limit order arrived and decreased when a market order arrived. However, the key problem with this line of work was that the generated mid-price return series exhibited a Hurst exponent of  $H \approx 0.25$  on all timescales. By contrast, as discussed in Section IV.H, real LOB data displays no long memory in mid-price returns (i.e.,  $H \approx 0.5$ ).

Challet and Stinchcombe (2001) refined Maslov's model by allowing multiple particles to be deposited on the pricing grid during a single timestep. They also allowed existing particles to evaporate, corresponding to the cancellation of an active order, although such evaporations were assumed to occur exogenously and independently for each particle, and independently of  $L(t)$ . In each time step, a random vector  $\gamma = (\gamma_1, \dots, \gamma_{A_n})$  was generated, where  $A_n$  was drawn from an exponential distribution then rounded to the nearest integer and  $\gamma_1, \dots, \gamma_{A_n}$  were independent draws from a normal distribution, again rounded to the nearest integer. An independent random vector  $\nu = (\nu_1, \dots, \nu_{B_m})$  was generated in the same way. For each  $\gamma_i$ ,  $i = 1, \dots, A_n$ , an  $A$  particle was deposited on the pricing grid with ask-relative price  $\gamma_i$ ; for each  $\nu_j$ ,  $j = 1, \dots, B_m$ , a  $B$  particle was deposited on the pricing grid with bid-relative price  $\nu_j$ . All  $A + B \rightarrow \emptyset$  annihilations occurred at prices with both  $A$  and  $B$  particles. Finally, all remaining particles evaporated independently with fixed probability, then the whole process was repeated. Challet and Stinchcombe's model exhibited a heavy-tailed return distribution and volatility clustering, and the Hurst exponent of the mid-price return series at large timescales was 0.5. The authors conjectured that it was the evaporations in their model, that had been absent in Maslov (2000), that made the Hurst exponent at large timescales match that of empirical data. However, at all shorter timescales Challet and Stinchcombe's model exhibited a Hurst exponent  $H < \frac{1}{2}$ , which is inconsistent with empirical data (as discussed in Section IV.H).

### 4. Continuous-time models

The generalization of zero-intelligence approaches to continuous time was due to a model first introduced by Daniels *et al.* (2002) and then developed by Smith *et al.*



(2003). Modelling market order arrivals and limit order arrivals and cancellations as independent Poisson processes, and modelling the relative price of an incoming limit order as being uniformly distributed on the semi-infinite interval  $(-s(t), \infty)$ , the evolution of the LOB was described by a master equation. The master equation was solved under a mean-field approximation that the depth available at neighbouring prices was independent, in the limit  $\delta p \rightarrow 0$  (i.e., assuming that the pricing grid was continuous, not discrete). Guided by dimensional analysis, the authors constructed simple, closed-form estimators for a variety of LOB properties, such as the mean spread, mean depth available at a given price, and mid-price diffusion, in terms of only the Poisson processes' arrival rates and the lot size  $\sigma$ . Similar results were achieved using Monte Carlo simulations. The model also provided possible explanations for why some empirical properties of LOBs varied across different markets (as discussed in Section IV). In particular, the lot size  $\sigma$  appeared explicitly in many of the closed-form estimators derived, and “phase transitions” between fundamentally different types of market behaviour were observed to occur as  $\sigma$  varied.

Many of the assumptions made by the Daniels *et al.* (2002) and Smith *et al.* (2003) model in order to maintain analytical tractability result in a poor resemblance to some aspects of real limit order markets. For example, in the limit  $\delta p \rightarrow 0$ , the only possible number of limit orders that can reside at a given price  $p$  is 0 or 1. This is because the relative price of an incoming limit order is chosen from the continuous uniform distribution, and therefore the probability of an incoming limit order having exactly the same price as an active order is infinitesimal. This destroyed the notion of limit orders “queueing up”, and thus removed what is a primary consideration for market participants: when to submit an order at the back of an existing priority queue versus when to start a new queue at a worse price (Parlour and Seppi, 2008). Additionally, by assuming that all the Poisson processes were independent of each other, the conditional structure of arrival rates known to exist in empirical data was discarded. But despite these simplifications, the model performed well when tested against some aspects of empirical data (Farmer *et al.*, 2005). Predictions of the mean spread  $\hat{s}$  and a measure of price diffusion  $d^{21}$  were made for 11 stocks from the LSE by calibrating the model's parameters using historical data. The empirical mean spread and price diffusion were then calculated directly from the data and compared these to the model's predictions us-

ing an ordinary least-squares regression. In particular, a straight line was fitted to the relationship

$$Z_{\text{emp}}(i) = zZ_{\text{mod}}(i) + c$$

where  $Z_{\text{emp}}(i)$  and  $Z_{\text{mod}}(i)$  were the mean empirical and model output values of statistic  $Z$  for stock  $i$ , for  $i = 1, \dots, 11$ . Under this set-up,  $z = 1$ ,  $c = 0$  would correspond to a perfect fit of the model to the data. For the mean spread, the ordinary least-squares estimates of the parameters were  $z = 0.99 \pm 0.10$  and  $c = 0.06 \pm 0.26$  and for the price diffusion, the ordinary least-squares estimates of the parameters were  $z = 1.33 \pm 0.10$  and  $c = 0.06 \pm 0.26$ . Bootstrap resampling was used to estimate the standard errors, as serial correlations within the data invalidated the assumptions required to make such estimation under ordinary least-squares regression. The model was also found to predict different price diffusions over different timescales, in agreement with empirical data. However, the distribution of price returns predicted by the model was thin-tailed on all timescales, thus poorly fitting empirical data.

Cont *et al.* (2010) recently introduced a variant of the Daniels *et al.* (2002) and Smith *et al.* (2003) model. Their main purpose was to study conditional (rather than equilibrium) behaviour: i.e., to understand how the frequency of occurrence of certain events was linked to  $L(t)$ . The model did not make the assumption that  $\delta p \rightarrow 0$ , thus ensuring that priority queues formed at discrete points on the price lattice. The assumption of a uniform distribution of relative prices (as made by Daniels *et al.* (2002) and Smith *et al.* (2003)) was also removed, and replaced by a power-law distribution, as suggested by empirical data (Bouchaud *et al.*, 2002; Cont *et al.*, 2010; Potters and Bouchaud, 2003; Zovko and Farmer, 2002).

Simulations of the Cont *et al.* (2010) model displayed the hump-shaped depth profile commonly reported in empirical data (see Section IV). Using Laplace transforms, the authors computed conditional probability distributions for several events, including an increase in  $m(t)$  at its next move, the matching of a limit order placed at  $b(t)$  before  $a(t)$  moved, and the matching of both a limit order at  $b(t)$  and a limit order at  $a(t)$  (“earning the spread”) before  $m(t)$  moved. Comparing the model's predictions to empirical market data for a single stock traded on the Tokyo Stock Exchange revealed fair, but not strong, agreement.

The Cont *et al.* (2010) model has recently been extended (Toke, 2011; Zhao, 2010) by revising the assumed arrival structure of market events. After conducting an empirical study of crude oil futures traded at the International Petroleum Exchange, Zhao (2010) rejected the assumption that the inter-arrival times of market events were independent draws from an exponential distribution, and thereby rejected the use of independent Poisson processes to model market event arrivals. Zhao replaced the independent Poisson processes in Cont *et al.*'s model

<sup>21</sup> Farmer *et al.* (2005) studied price diffusion by calculating the variance  $v_\tau$  of the set  $\{(m(t_i + \tau) - m(t_i)) \mid i = 1, \dots, n\}$  for various values of  $\tau$ , where  $\{t_i \mid i = 1, \dots, n\}$  was the set of times at which the mid-price changed. They then performed a least-squares regression to estimate  $d$  in the expression  $v_\tau = d\tau$ .

with a Hawkes process<sup>22</sup> (Bauwens and Hautsch, 2009) that described the arrival rate of all market events as a function of recent order arrival rates and of the number of arrivals that had recently occurred. When an arrival occurred, its type (e.g., market order arrival, limit order cancellation) was determined exogenously. This produced order flows in which periods of high arrival rates clustered together in time, and periods of low arrival rates clustered together in time, in agreement with empirical data (Ellul *et al.*, 2003; Hall and Hautsch, 2006). Zhao demonstrated that this modification to the model of Cont *et al.* resulted in an improved fit of the empirically observed mean relative depth profile. Toke (2011) similarly replaced the Poisson processes in Cont *et al.*'s model with Hawkes processes, but unlike Zhao, Toke used multiple mutually-exciting Hawkes processes, one for each type of market event. By studying empirical data from several different asset classes, Toke observed that once a market order had been placed, the mean time until the next limit order was placed was less than the corresponding unconditional mean time. The use of Hawkes processes allowed a coupling between the arrival rates of limit orders and market orders, which produced simulated order flow that matched his empirical observations. The distribution of spreads generated by the Hawkes processes was closer to the true distribution of spreads than that generated by a Poisson process model.

Cont and de Larrard (2011) recently introduced a model in which only  $n(b(t), t)$  and  $n(a(t), t)$  were tracked, rather than the whole depth profile. When the depth available at either of the prices became zero, they assumed that the depth available at the next best price was a random variable drawn from a distribution  $f$ . The state space of this model is  $\mathbb{N}^2$ , rather than  $\mathbb{Z}^P$  as in most other recent LOB models. The justification for such a simplified set-up was that, in real time, many market participants only have access to the depths available at the best prices, not the whole depth profile (although this is increasingly less common as electronic trading platforms deliver ever more up-to-date information about the LOB in real time to market participants (Boehmer *et al.*, 2005; Bortoli *et al.*, 2006)). Market order arrivals, limit order arrivals, and limit order cancellations were assumed to be governed by independent Poisson processes. Analytical estimates for several market properties – such as volatility, the distribution of time until the next change in  $m(t)$ , the distribution and autocorrelation of price changes, and the conditional probability that  $m(t)$  moves in a specified direction given the depths available at  $b(t)$  and  $a(t)$  – were deduced solely in terms of the Poisson processes'

rate parameters and the distribution  $f$ . Additionally, different levels of autocorrelation of the mid-price series were shown to emerge from the model at different sampling frequencies, in agreement with empirical observations (Cont, 2001; Zhou, 1996).

## 5. Beyond zero intelligence

The models discussed so far in this section all attempt to maintain analytical tractability in a zero-intelligence framework. Mike and Farmer (2008) took a different approach, and instead designed a zero-intelligence model that mimicked very closely the order flows they observed in an empirical study of data from the LSE. The model assumed that the relative price of each incoming order was drawn independently from a Student's  $t$  distribution, and closely matched cancellation rates for active orders to empirical data.

For stocks with a small tick size and low volatility, the model was found to correctly exhibit negative autocorrelation of logarithmic mid-price returns on short timescales. Furthermore, the model was found to make good predictions of the distribution of mid-price returns, including heavy tails, and of  $s(t)$ . This is a substantial improvement on previous models, whose predictions concerned only the mean values of such statistics rather than their whole distribution. However, for other stocks, the model was found to perform less well.

Although Mike and Farmer (2008) did not assume that any market participants are rational, nor that they are attempting to maximize some personal utility by trading, the highly conditional structure of random variables in their model suggest ways in which observed regularities in order flow might be motivated by rational decision-making. For example, the existence of a higher cancellation rates near  $b(t)$  and  $a(t)$  can be interpreted as the impatience felt by market participants who choose to submit limit orders at such aggressive prices in the first place. The lower rate of cancellation deeper inside the LOB reflects the fact that market participants would not submit such orders unless they were willing to wait patiently for them to be filled at some point in the future.

Gu and Zhou (2009a) simulated the Mike and Farmer (2008) model and performed a DFA $m$  (see Section III.M) on the output mid-price return and volatility series. They found that neither series exhibited a long memory. For the mid-price return series, this is an accurate replication of the stylized facts (see Section IV.H); whereas for the volatility series, this indicates that the model fails to capture the empirically observed long memory. Gu and Zhou then proposed an extension to the model, in which the relative prices were drawn from a Student's  $t$  distribution with long memory, rather than independently. They found that such a modification caused the volatility series to exhibit long memory, in line with the stylized

<sup>22</sup> A Hawkes process is a point process with time-varying intensity parameter  $\lambda(t) = \lambda_0(t) + \sum_{t_i < t} \sum_j C_j e^{-D_j(t-t_i)}$ , where the  $t_i$  denote the times of previous arrivals and the  $C_j$  and  $D_j$  are parameters controlling the intensity of arrivals.

facts, while still retaining all of the model's other results.

Gu and Zhou (2009b) replaced the Student's  $t$  distribution for relative prices of incoming orders with several other distributions, and simulated the Mike and Farmer (2008) model to examine how this affected its output. They found that the empirically observed power-law tail in the mid-price return distribution only appeared in the model's output when the distribution from which positive relative prices were drawn had heavy tails, irrespective of whether the distribution from which negative relative prices were drawn had heavy tails or not.

### C. Agent-based models

An agent-based model is a model in which a large number of (possibly heterogeneous) agents, each with specific rules governing their behaviour, are assumed to interact in a specified way. Both the performance of individual agents and the aggregate effect of all agents in the system can be studied, either analytically, or via Monte Carlo simulation. By allowing each individual agent's behaviour to be specified without any explicit requirements regarding rationality in a given circumstance, agent-based models lie between the two extremes of zero-intelligence and perfect-rationality models.

Chakraborti *et al.* (2011a) highlighted that a key advantage of agent-based models in comparison to zero-intelligence and perfect-rationality models that aggregate across market participants is that heterogeneity between different market participants in real markets can be incorporated directly. Models with independent and identically distributed order flows can then be considered as special cases of agent-based models in which there is a single representative agent in the market and of which all other market participants are a perfect clone.

However, agent-based models of limit order markets also have drawbacks. Due to the large number of interacting components in a LOB, agent-based modelling makes it difficult to track explicitly how a specified input parameter affects the output. It is also very difficult to encode quantitatively the complex and interacting strategies of market participants in a real limit order market into a set of rules governing an agent in an agent-based model, and finding a set of agent rules that produces a specific behaviour from the model provides no guarantee that such a set of rules is the only one to do so (Preis *et al.*, 2007). Abergel and Jedidi (2011) attempted to address these issues by explicitly deriving systems of stochastic differential equations that described the order flows arising from specified agent-based models. Such equations could then be studied analytically, and equations for the price dynamics were derived in terms of the agent-based model's input parameters, thereby demonstrating the exact links between the two. For example, a very simple model was shown to result in Gaussian pro-

cess dynamics, with a diffusion coefficient that depended on the model's input parameters. Volatility dynamics were similarly derived.

Early agent-based models of LOBs assumed that agents arrived sequentially at the market (Foucault *et al.*, 2005), and that the LOB emptied at the end of every time step. Such a set-up failed to acknowledge the LOB's key functionality of storing supply and demand for later consumption by other market participants (Smith *et al.*, 2003). However, more recent agent-based models have more closely mimicked the process of real limit order trading, and have been able to reproduce a wide range of empirical features present in LOB data (Challet and Stinchcombe, 2003; Chiarella and Iori, 2002; Cont and Bouchaud, 2000; Preis *et al.*, 2006).

Cont and Bouchaud (2000) showed that when agents were assumed to imitate each other, a heavy-tailed return distribution emerged. The model's output also exhibited clustered volatility and aggregational Gaussianity, as discussed in Section IV.H.

Chiarella and Iori (2002) noted that if all agents were assumed to share a common valuation regime for the asset being traded, the realized volatility was too low compared to empirical data and there was no volatility clustering. They thereby argued that substantial heterogeneity must exist between market participants in order for the highly non-trivial properties of volatility, as discussed in Section III.H, to emerge in real limit order markets. Cont (2005) noted that differences in agents' timescales (i.e., their level of impatience) could be a source of such heterogeneity in real markets.

Preis *et al.* (2006) reproduced the main findings of the Smith *et al.* (2003) model, but using an agent-based model rather than independent Poisson processes. By finely tuning agents' trading strategies, the heavy-tailed distribution of mid-price returns, the super-diffusivity of mid-price returns over medium timescales, and the negative autocorrelation of  $m(t)$  on an event-by-event timescale were all reproduced by the model. The performance of individual agents in the model has also been studied (Preis *et al.*, 2007). The Hurst exponent  $H$  of the mid-price series was found to vary according to the number of agents in the model. The best fit of  $H$  against values calculated from empirical data was found to occur with 150 to 500 "liquidity provider" (i.e., limit order placing) agents and 150 to 500 "liquidity taker" (i.e., market order placing) agents in the model.

Challet and Stinchcombe (2003) highlighted that many early LOB models assumed that model parameters (such as the arrival rate of new limit orders) were constant through time. They therefore produced an agent-based model with parameters that varied over time, and compared its output to a version of the same model where such parameters were instead held fixed. They found that allowing the parameters to vary resulted in the emergence of a heavy-tailed distribution of mid-price movements,

autocorrelated mid-price returns, and volatility clustering.

Lillo (2007) showed how an agent-based model could explain the empirically observed power-law distribution of relative prices of incoming orders (see Section IV.B). In particular, he solved a utility maximization problem to show that if mid-price movements were assumed to followed a Brownian motion, each perfectly rational agent should choose the relative price of their submitted orders to be:

$$\Delta_x^* = \sqrt{2}g^{-1}(\alpha)VT^{\frac{1}{2}} \quad (9)$$

where  $g(\alpha)$  describes the individual agent's risk aversion,  $T$  is the individual agent's maximum time horizon (i.e., the maximum length of time that the agent is willing to wait before performing the trade), and  $V$  is the market volatility. He then studied how empirically observed homogeneity in  $g$  and  $T$  and empirically observed fluctuations in  $V$  affected the relative price choices of many interacting agents with different risk aversions  $g$  and different maximum time horizons  $T$ . He found that heterogeneity in  $T$  was the most likely source of the power-law tails in the distribution of  $\Delta_x$ , and that the homogeneity in  $g$  and fluctuations in  $V$  that were empirically observed in a wide range of markets were unlikely to lead to a power-law tail in the distribution of  $\Delta_x$ .

## VI. DISCUSSION

To date, no one has formulated a market mechanism design problem to which the LOB is known to be the optimal solution (Parlour and Seppi, 2008). However, by giving every market participant the freedom to evaluate their own need for immediate liquidity, LOBs have revolutionized the process of trading. The rewards (and, indeed, the risks) associated with patience are now available to all, rather than being reserved for a small number of market makers. New trading strategies have emerged from limit order trading (Bertsimas and Lo, 1998; Biais *et al.*, 1995; Obizhaeva and Wang, 2005), and market participants must now continually consider the trade-off between submitting orders that match at a better price and submitting orders that match in a shorter time.

Empirical studies and theoretical models have deepened understanding of specific aspects of limit order trading. However, as we have discussed at length, a key unresolved question is how the various pieces of the puzzle fit together. For example, models that accurately capture the dynamics of the price process on an event-by-event timescale poorly reproduce price dynamics on inter-day timescales. Similarly, models that explain price dynamics on inter-day timescales offer little understanding of how such dynamics are motivated by the LOB microstructure.

As discussed in Section IV.H, several stylized facts have been empirically observed in a wide range of markets.

Many authors (e.g., (Gu and Zhou, 2009a; Lillo, 2007; Stanley *et al.*, 2008)) agree that one of the main challenges facing researchers of LOBs today is to understand these stylized facts better. LOB models can help us to achieve this, and some progress has been made. However, no single model has yet been capable of simultaneously reproducing all of the stylized facts, nor is there a clear picture about precisely how the stylized facts emerge as a consequence of the actions of many heterogeneous market participants. This continues to be an active area of research.

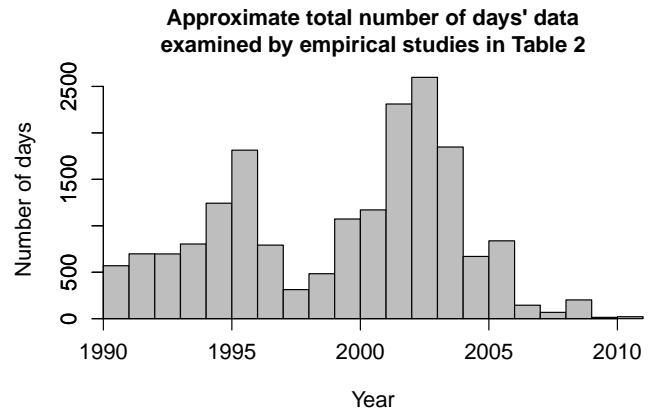


FIG. 9 Number of total days each year examined by empirical studies listed in Appendix A

A great deal of effort has been invested in the study of empirical LOB data. Figure 9 displays a plot of the approximate number of days' data per year that studies discussed in this article have examined, as described in Appendix A. Although the breadth of such empirical work is substantial, the overwhelming picture painted by Figure 9 is that the data studied is old. Often, it is also of poor quality. Strong assertions have been made in empirical studies based on single stocks over very short time periods. The data studied rarely includes all order flows at all prices, so extensive auxiliary assumptions are often required before any statistical analysis can even begin. Additionally, markets change over time, so empirical observations from more than a decade ago might not accurately describe current LOB activity.

There are also substantial challenges associated with studying historical limit order book data. Several LOB properties are believed to exhibit a long memory, although directly testing this hypothesis on a finite data set is a difficult task. Estimating the Hurst exponent  $H$  is similarly problematic, and several empirical studies contain systematic errors in their calculations. This makes quantitative comparisons between potential long-memory processes very difficult, and it is unclear whether the differences in the reported estimates of Hurst exponents are really a result of differences in different markets,

or simply a result of differences in methodology when performing such estimation. Furthermore, such long-range correlations make it difficult to estimate confidence intervals for several LOB statistics, as the effective sample size of observations is far smaller than the number of data points. Studies of recent, high-quality LOB data that are conducted with stringent awareness of these potential statistical pitfalls are needed to understand better the LOBs of today.

Direct comparisons between different empirical studies are also problematic. The behaviour and strategies of market participants may have changed over time, meaning that the date over which data was collected might itself play a role in the statistics observed within it. This is particularly important with the recent surge in popularity of electronic trading algorithms, which are able to process data to interpret changing market conditions, then to submit or cancel orders accordingly, in a fraction of the time that it would take a human to perform the same task. However, the date range of a study is far from being the only factor determining the statistical behaviour of empirical data. The sampling frequency, asset class studied, LOB resolution parameters, specific trade-matching nuances, and many other factors all influence empirical data, but all vary from study to study. This makes it difficult to address questions about how altering any one of these factors in isolation might change the observed statistics.

Another aspect that is clear from empirical studies is how poorly the data supports the very strong assumptions made by many LOB models. Although every model must make assumptions to facilitate computation, many LOB models have been built on elaborate and inaccurate assumptions that make it almost impossible to relate their output to real LOBs. Indeed, this is a common problem when using purely statistical models to probe LOB data, as it is often unclear how aspects of the models relate to specific elements of a LOB.

Although precisely what is meant by “equilibrium” depends upon context, almost all LOB models to date have focused on some form of equilibrium, such as a Markov-perfect equilibrium in sequential-game models or a state-space equilibrium in reaction-diffusion models. However, overwhelming empirical evidence suggests that LOBs are subjected to constant shocks and therefore always display transient behaviour. Early work on models that are not in equilibrium has hinted at promising results (Challet and Stinchcombe, 2003), but more work in this area is needed.

Both perfect-rationality and zero-intelligence approaches to LOB modelling have proven useful in enabling analytic computation and numerical simulation, yet both require strong assumptions that are not justified by data. Agent-based models appear to offer some compromise between the two extremes. They can operate within a zero-intelligence framework (if no ratio-

nal decision-making is programmed into agents’ specifications); a perfectly-rational framework (if rational decision-making based on the maximization of a utility function is the only specification driving agents’ behaviour); or, crucially, they can reside somewhere between. Furthermore, the level of game-theoretic considerations involved in agents’ decision-making can also be controlled by specifying how strongly agents react to each other and forecast each other’s actions. Therefore, agent-based models have the potential to provide a rich toolbox for investigating LOBs without a requirement for extreme modelling assumptions. However, it remains unclear whether agent-based models of limit order trading really offer a deeper understanding of market dynamics or merely amount to curve-fitting exercises in which parameters are varied until some form of non-trivial behaviour emerges. Nevertheless, recent developments suggest that the performance of LOB models can be improved by removing the inherent homogeneity associated with many zero-intelligence approaches (Toke, 2011; Zhao, 2010). Consequently, the heterogeneity offered by agent-based models might pave the way for new explanations of LOB phenomena.

Price changes and volatility are among the most hotly debated topics in limit order markets (Almgren and Chriss, 2001; Bouchaud *et al.*, 2009; Hasbrouck, 1991; Potters and Bouchaud, 2003). What causes volatility to vary over time? Why should periods of high activity cluster together in time? Why should price fluctuations be so frequent (and so large) on intra-day timescales, given that external news events occur so rarely (Maslov, 2000)? It is not even agreed whether the number of market orders (Jones *et al.*, 1994), the size of market orders (Gallant *et al.*, 1992), or liquidity fluctuations in the LOB (Bouchaud *et al.*, 2009) play the dominant role in determining volatility. It seems likely that the answers to such questions will not be found in isolation, but rather that there is an intricate interplay between the many parts of the “volatility puzzle”. Recent work has attempted to tie some of these ideas together. Specifically, Bouchaud *et al.* (2009) conjectured that volatility might be better understood by considering the need for market participants to minimize their market impact. In particular, although news events happen rarely, when they do occur they cause market participants to want to buy or sell very large quantities of the asset being traded. Market participants understand that they cannot simply perform a large trade immediately, as the market impact of their actions would cause them to trade at very unfavourable prices. Instead, they break up large trades into smaller chunks that are then gradually processed over weeks or even months. Due to their differing needs for immediacy, or indeed their differing reaction to the released news, different market participants choose different times for the submission of such chunks, causing a cascading of the original news event to the submission of multiple dif-

ferent orders at multiple different times. It will be interesting to observe whether this explanation withstands closer examination.

Price impact and market impact also continue to be active areas of research. Indeed, a deeper understanding of these notions is very desirable, as they form a conceptual bridge between the microstructure mechanics of order matchings and the macroeconomic concepts of price formation. Considerations about price impact and market impact could also help to explain the actions of market participants in certain situations. For example, Wyart *et al.* (2008) conjectured that the empirically observed cross-correlation between volatility and the relative price of incoming limit orders might be a result of market participants carefully managing their market impact. Gatheral (2010) has shown that if the instantaneous mid-price impact function is nonlinear in market order size  $\omega_x$  – and the empirical evidence certainly suggests this is the case (Hasbrouck, 1991; Lillo *et al.*, 2003; Potters and Bouchaud, 2003) – then it is possible to deduce bounds on the way that the LOB must repopulate if arbitrage opportunities are to be excluded. However, despite the striking regularities that have emerged from empirical studies, little is understood about *why* the functional forms of price impact functions are what they are, and almost nothing is understood about market impact. Additionally,  $L(t)$  clearly plays an important role in determining the price and market impacts of an action, and it has recently been observed (Cont *et al.*, 2011) that tracking only the mean impact of individual market orders might be insufficient to gain clear insight into price impact and market impact. To remedy this problem, it has been proposed that price impact should be studied not only as a function of arriving order size (or imbalance) but also as a function of  $L(t)$  at the time of order submissions (Cristelli *et al.*, 2010). However, the curse of dimensionality poses substantial problems, as the state space of  $L(t)$  is so large.

As ever more electronic limit order trading platforms have become available, it has become increasingly common for specific assets to be traded on multiple electronic LOBs simultaneously. This poses a problem for empirical research, as the study of any individual LOB in isolation no longer provides a snapshot of the “whole” limit order market for the asset. Furthermore, differences in matching rules and transaction costs across different trading platforms make it difficult to directly compare different LOBs. Encouragingly, recent work has found similar behaviours when studying different LOBs that traded the same asset simultaneously (Cont *et al.*, 2011), but there is no reason to assume that this will always be the case. It is beneficial for market participants to have the option of trading the same asset on multiple platforms, as competition between different exchanges drives technological innovation and reduces the amount of market downtime. However, understanding how to assimilate data across

multiple platforms will be of paramount importance in future studies.

Finally, in addition to being a popular trade-matching algorithm that offers market participants greater choice and flexibility than ever before, LOBs are a rich and exciting testing ground for theories. Both empirical data and LOB models have provided new insight into longstanding economic questions such as market efficiency, price formation, and the rationality of market participants. Furthermore, LOBs are a classic example of a complex system. Despite the deceptively simple rules governing trade, several hallmarks of complexity, including nonlinear feedback, scaling, and universality, are present in LOBs. Both the quantity and the quality of LOB data that is available far exceeds that of many other studied complex systems. It will be interesting to see what new insights into not just trading, but complex systems as a whole, the study of LOBs is able to provide in the future.

## ACKNOWLEDGMENTS

We would like to thank Bruno Biais, Jean-Philippe Bouchaud, J. Dooyne Farmer, Gabriele La Spada, Sergei Maslov, Stephen Roberts, Torsten Schöneborn, Cosma Shalizi, Neil Shephard, D. Eric Smith, Jonathan Tse, Thaleia Zariphopoulou, and Wei-Xing Zhou for useful discussions. MDG would like to thank EPSRC (Industrial CASE Award 08001834), HSBC Bank, and the Oxford-Man Institute of Quantitative Finance for supporting this work.

## REFERENCES

- Abergel, F., and A. Jedidi (2011), “A mathematical approach to order book modelling,” in *Econophysics of Order-driven Markets: Proceedings of Econophys-Kolkata V*, edited by F. Abergel, B. K. Chakrabarti, C. A., and M. M. (Springer, Milan) pp. 93–107.
- Aït-Sahalia, Y., P. Mykland, and L. Zhang (2005), “Ultra high frequency volatility estimation with dependent microstructure noise,” University of Chicago Preprint, available at [galton.uchicago.edu/~mykland/paperlinks/depnoise.pdf](http://galton.uchicago.edu/~mykland/paperlinks/depnoise.pdf).
- Alfonsi, A., A. Fruth, and A. Schied (2010), “Optimal execution strategies in limit order books with general shape functions,” *Quantitative Finance* **10** (2), 143.
- Almgren, R., and N. Chriss (2001), “Optimal execution of portfolio transactions,” *Journal of Risk* **3** (2), 5.
- Anand, A., S. Chakravarty, and T. Martell (2005), “Empirical evidence on the evolution of liquidity: choice of market versus limit orders by informed and uninformed traders,” *Journal of Financial Markets* **8** (3), 288.
- Andersen, T. G., and V. Todorov (2010), “Realized volatility and multipower variation,” in *Encyclopedia of Quantitative Finance*, edited by R. Cont (Wiley) pp. 1494–1500.
- Bak, P., M. Paczuski, and M. Shubik (1997), “Price variations

- in a stock market with many agents,” *Physica A* **246** (3-4), 430.
- Bandi, F., and J. Russell (2006), “Separating microstructure noise from volatility,” *Journal of Financial Economics* **79** (3), 655.
- Barndorff-Nielsen, O. E., and N. Shephard (2010), “Volatility,” in *Encyclopedia of Quantitative Finance*, edited by R. Cont (Wiley) pp. 1898–1901.
- Bauwens, L., and N. Hautsch (2009), “Modelling financial high frequency data using point processes,” in *Handbook of Financial Time Series*, edited by T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Springer, Berlin) pp. 953–979.
- Beran, J. (1994), *Statistics for long-memory processes*, Vol. 61 (Chapman & Hall).
- Bertsimas, D., and A. Lo (1998), “Optimal control of execution costs,” *Journal of Financial Markets* **1** (1), 1.
- Biais, B., P. Hillion, and C. Spatt (1995), “An empirical analysis of the limit order book and the order flow in the Paris Bourse,” *The Journal of Finance* **50** (5), 1655.
- Biais, B., P. Hillion, and C. Spatt (1999), “Price discovery and learning during the preopening period in the Paris Bourse,” *Journal of Political Economy* **107** (6), 1218.
- Boehmer, E., G. Saar, and L. Yu (2005), “Lifting the veil: an analysis of pre-trade transparency at the NYSE,” *The Journal of Finance* **60** (2), 783.
- Bortoli, L., A. Frino, E. Jarnecic, and D. Johnstone (2006), “Limit order book transparency, execution risk, and market liquidity: evidence from the Sydney Futures Exchange,” *Journal of Futures Markets* **26** (12), 1147.
- Bouchaud, J. P., J. D. Farmer, and F. Lillo (2009), “How markets slowly digest changes in supply and demand,” (North-Holland, San Diego) pp. 57–160.
- Bouchaud, J. P., Y. Gefen, M. Potters, and M. Wyart (2004), “Fluctuations and response in financial markets: the subtle nature of ‘random’ price changes,” *Quantitative Finance* **4** (2), 176.
- Bouchaud, J. P., M. Mézard, and M. Potters (2002), “Statistical properties of stock order books: empirical results and models,” *Quantitative Finance* **2** (4), 251.
- Bouchaud, J. P., and M. Potters (2003), *Theory of financial risk and derivative pricing: from statistical physics to risk management* (Cambridge University Press).
- Cao, C., O. Hansch, and X. Wang (2008), “Order placement strategies in a pure limit order book market,” *Journal of Financial Research* **31** (2), 113.
- Carrie, C. (2006), “The new electronic trading regime of dark books, mashups and algorithmic trading,” in *Algorithmic Trading II: Precision, Control, Execution*, edited by B. R. Bruce (Institutional Investor Journals) pp. 14–20.
- Chakraborti, A., I. M. Toke, M. Patriarca, and F. Abergel (2011a), “Econophysics: agent-based models,” *Quantitative Finance* **11** (7), 1013.
- Chakraborti, A., I. M. Toke, M. Patriarca, and F. Abergel (2011b), “Econophysics: empirical facts,” *Quantitative Finance* **11** (7), 991.
- Chakravarty, S., and C. W. Holden (1995), “An integrated model of market and limit orders,” *Journal of Financial Intermediation* **4** (3), 213.
- Challet, D., and R. Stinchcombe (2001), “Analyzing and modeling 1 + 1d markets,” *Physica A* **300** (1-2), 285.
- Challet, D., and R. Stinchcombe (2003), “Non-constant rates and over-diffusive prices in a simple model of limit order markets,” *Quantitative Finance* **3** (3), 155.
- Chan, D. L. C., D. Eliezer, and I. I. Kogan (2001), “Numerical analysis of the minimal and two-liquid models of the market microstructure,” arXiv:0101474.
- Chiarella, C., and G. Iori (2002), “A simulation analysis of the microstructure of double auction markets,” *Quantitative Finance* **2** (5), 346.
- Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009), “Power-law distributions in empirical data,” *SIAM Review* **51** (4), 661.
- Cont, R. (2001), “Empirical properties of asset returns: stylized facts and statistical issues,” *Quantitative Finance* **1** (2), 223.
- Cont, R. (2005), “Long range dependence in financial markets,” in *Fractals in Engineering*, edited by J. Lévy-Véhel and E. Lutton (Springer, London).
- Cont, R., and J. Bouchaud (2000), “Herd behavior and aggregate fluctuations in financial markets,” *Macroeconomic Dynamics* **4** (2), 170.
- Cont, R., A. Kukanov, and S. Stoikov (2011), “The price impact of order book events,” arXiv:1011.6402.
- Cont, R., and A. de Larrard (2011), “Price dynamics in a Markovian limit order market,” arXiv:1104.4596.
- Cont, R., M. Potters, and J. P. Bouchaud (1997), “Scaling in stock market data: stable laws and beyond,” in *Scale invariance and beyond: Les Houches Workshop, March 10-14, 1997*, edited by B. Dubrulle, F. Graner, and D. Sornette (Springer).
- Cont, R., S. Stoikov, and R. Talreja (2010), “A stochastic model for order book dynamics,” *Operations Research* **58** (3), 549.
- Copeland, T. E., and D. Galai (1983), “Information effects on the bid-ask spread,” *Journal of Finance* **38** (5), 1457.
- Cristelli, M., V. Alfi, L. Pietronero, and A. Zaccaria (2010), “Liquidity crisis, granularity of the order book and price fluctuations,” *The European Physical Journal B* **73** (1), 41.
- Daniels, M. G., J. D. Farmer, L. Gillemot, G. Iori, and E. Smith (2002), “A quantitative model of trading and price formation in financial markets,” arXiv:0112422.
- Drożdż, S., M. Forczek, J. Kwapien, P. Oświęcimka, and R. Rak (2007), “Stock market return distributions: From past to present,” *Physica A* **383** (1), 59.
- Dufour, A., and R. F. Engle (2000), “Time and the price impact of a trade,” *The Journal of Finance* **55** (6), 2467.
- Eisler, Z., J. P. Bouchaud, and J. Kockelkoren (2010), “The price impact of order book events: market orders, limit orders and cancellations,” arXiv:0904.0900.
- Eliezer, D., and I. I. Kogan (1998), “Scaling laws for the market microstructure of the interdealer broker markets,” arXiv:9808240.
- Ellul, A., C. W. Holden, P. Jain, and R. Jennings (2003), “Determinants of order choice on the New York Stock Exchange,” *Indiana University Preprint*, available at <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.114.6359>.
- Engle, R. F., and A. J. Patton (2004), “Impacts of trades in an error-correction model of quote prices,” *Journal of Financial Markets* **7** (1), 1.
- Evans, M. D. D., and R. K. Lyons (2002), “Order flow and exchange rate dynamics,” *Journal of Political Economy* **110** (1), 170.
- Farmer, J. D., and D. Foley (2009), “The economy needs agent-based modelling,” *Nature* **460** (7256), 685.
- Farmer, J. D., and F. Lillo (2004), “On the origin of power-

- law tails in price fluctuations,” *Quantitative Finance* **4** (1), 7.
- Farmer, J. D., P. Patelli, and I. I. Zovko (2005), “The predictive power of zero intelligence in financial markets,” *Proceedings of the National Academy of Sciences of the United States of America* **102** (6), 2254.
- Field, J., and J. Large (2008), “Pro-rata matching and one-tick futures markets,” University of Oxford Preprint, available at <http://www.economics.ox.ac.uk/members/jeremy.large/ProRataApril08.pdf>.
- Foucault, T. (1999), “Order flow composition and trading costs in a dynamic limit order market,” *Journal of Financial Markets* **2** (2), 99.
- Foucault, T., O. Kadan, and E. Kandel (2005), “Limit order book as a market for liquidity,” *Review of Financial Studies* **18** (4), 1171.
- Friedman, D. (2005), “The double auction market institution: a survey,” in *The Double Auction Market: Institutions, Theories, and Evidence*, edited by D. Friedman and J. Rust (Addison-Wesley).
- Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley (2006), “Institutional investors and stock market volatility,” *Quarterly Journal of Economics* **121** (2), 461.
- Gallant, A. R., P. E. Rossi, and G. Tauchen (1992), “Stock prices and volume,” *Review of Financial Studies* **5** (2), 199.
- Gatheral, J. (2010), “No-dynamic-arbitrage and market impact,” *Quantitative Finance* **10** (7), 749.
- Geweke, J., and S. Porter-Hudak (1983), “The estimation and application of long-memory time-series models,” *Journal of Time Series Analysis* **4** (4), 221.
- Glosten, L. R. (1994), “Is the electronic open limit order book inevitable?” *The Journal of Finance* **49** (4), 1127.
- Glosten, L. R., and P. R. Milgrom (1985), “Bid, ask and transaction prices in a specialist market with heterogeneously informed traders,” *Journal of Financial Economics* **14** (1), 71.
- Gode, D. K., and S. Sunder (1993), “Allocative efficiency of markets with zero-intelligence traders: market as a partial substitute for individual rationality,” *Journal of Political Economy* **101** (1), 119.
- Goettler, R., C. Parlour, and U. Rajan (2006), “Microstructure effects and asset pricing,” Preprint, available at <http://en.scientificcommons.org/33345856>.
- Gopikrishnan, P., M. Meyer, L. A. N. Amaral, and H. E. Stanley (1998), “Inverse cubic law for the distribution of stock price variations,” *The European Physical Journal B-Condensed Matter and Complex Systems* **3** (2), 139.
- Gopikrishnan, P., V. Plerou, L. A. N. Amaral, M. Meyer, and H. E. Stanley (1999), “Scaling of the distribution of fluctuations of financial market indices,” *Physical Review E* **60** (5), 5305.
- Gopikrishnan, P., V. Plerou, X. Gabaix, and H. E. Stanley (2000), “Statistical properties of share volume traded in financial markets,” *Physical Review E* **62** (4), 4493.
- Gould, M. D., M. A. Porter, S. Williams, M. McDonald, D. J. Fenn, and S. D. Howison (2011), “Statistical properties of foreign exchange limit order books,” Working paper.
- Grossman, S. J., and J. E. Stiglitz (1980), “On the impossibility of informationally efficient markets,” *The American Economic Review* **70** (3), 393.
- Gu, G. F., W. Chen, and W. X. Zhou (2008a), “Empirical distributions of chinese stock returns at different microscopic timescales,” *Physica A* **387**, 495.
- Gu, G. F., W. Chen, and W. X. Zhou (2008b), “Empirical regularities of order placement in the chinese stock market,” *Physica A* **387** (13), 3173.
- Gu, G. F., W. Chen, and W. X. Zhou (2008c), “Empirical shape function of limit-order books in the chinese stock market,” *Physica A* **387** (21), 5182.
- Gu, G. F., and W. X. Zhou (2009a), “Emergence of long memory in stock volatility from a modified Mike-Farmer model,” *Europhysics Letters* **86**, 48002.
- Gu, G. F., and W. X. Zhou (2009b), “On the probability distribution of stock returns in the mike-farmer model,” *European Physical Journal B* **67** (4), 585.
- Guillaume, D. M., M. M. Dacorogna, R. R. Davé, U. A. Müller, R. B. Olsen, and O. V. Pictet (1997), “From the bird’s eye to the microscope: a survey of new stylized facts of the intra-daily foreign exchange markets,” *Finance and Stochastics* **1** (2), 95.
- Hall, A. D., and N. Hautsch (2006), “Order aggressiveness and order book dynamics,” *Empirical Economics* **30** (4), 973.
- Hamilton, J. D. (1994), *Time series analysis*, Vol. 2 (Cambridge University Press).
- Harris, L. (2003), *Trading and exchanges: Market microstructure for practitioners* (Oxford University Press).
- Harris, L., and J. Hasbrouck (1996), “Market vs. limit orders: the SuperDOT evidence on order submission strategy,” *Journal of Financial and Quantitative Analysis* **31** (2), 213.
- Hasbrouck, J. (1991), “Measuring the information content of stock trades,” *Journal of Finance* **46** (1), 179.
- Hasbrouck, J., and G. Saar (2002), “Limit orders and volatility in a hybrid market: The Island ECN,” NYU Working Paper No. FIN-01-025, available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1294561](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1294561).
- Hautsch, N., and R. Huang (2009), “The market impact of a limit order,” Humboldt Universität Preprint, available at <http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2009-051.pdf>.
- Hendershott, T., and C. M. Jones (2005), “Island goes dark: transparency, fragmentation, and regulation,” *Review of Financial Studies* **18** (3), 743.
- Hendershott, T., C. M. Jones, and A. J. Menkveld (2011), “Does algorithmic trading improve liquidity?” *The Journal of Finance* **66** (1), 1.
- Hill, B. M. (1975), “A simple general approach to inference about the tail of a distribution,” *The Annals of Statistics*, 1163.
- Hollifield, B., R. A. Miller, and P. Sandås (2004), “Empirical analysis of limit order markets,” *The Review of Economic Studies* **71** (4), 1027.
- Hollifield, B., R. A. Miller, P. Sandås, and J. Slive (2006), “Estimating the gains from trade in limit-order markets,” *The Journal of Finance* **61** (6), 2753.
- Jones, C. M., G. Kaul, and M. L. Lipson (1994), “Transactions, volume, and volatility,” *Review of Financial Studies* **7** (4), 631.
- Kantelhardt, J. W., E. Koscielny-Bunde, H. H. A. Rego, S. Havlin, and A. Bunde (2001), “Detecting long-range correlations with detrended fluctuation analysis,” *Physica A* **295** (3), 441.
- Kempf, A., and O. Korn (1999), “Market depth and order size,” *Journal of Financial Markets* **2** (1), 29.
- Kyle, A. S. (1985), “Continuous auctions and insider trading,” *Econometrica* **53** (6), 1315.
- La Spada, G., and F. Lillo (2011), “The effect of round-off error on long memory processes,” arXiv:1107.4476.



- Lillo, F. (2007), "Limit order placement as an utility maximization problem and the origin of power law distribution of limit order prices," *European Physical Journal B* **55** (4), 453.
- Lillo, F., and J. D. Farmer (2004), "The long memory of the efficient market," *Studies in Nonlinear Dynamics and Econometrics* **8** (3), 1.
- Lillo, F., J. D. Farmer, and R. N. Mantegna (2003), "Econophysics: master curve for price-impact function," *Nature* **421** (6919), 129.
- Lillo, F., S. Mike, and J. D. Farmer (2005), "Theory for long memory in supply and demand," *Physical Review E* **71** (6), 066122.
- Liu, Y., P. Cizeau, M. Meyer, C. K. Peng, and H. Eugene Stanley (1997), "Correlations in economic time series," *Physica A* **245** (3-4), 437.
- Liu, Y., P. Gopikrishnan, P. Cizeau, M. Meyer, C. K. Peng, and H. E. Stanley (1999), "Statistical properties of the volatility of price fluctuations," *Physical Review E* **60** (2), 1390.
- Lo, A. W. (1989), *Long-term memory in stock market prices*, Tech. Rep. (National Bureau of Economic Research).
- Lo, A. W., and A. C. MacKinlay (2001), *A non-random walk down Wall Street* (Princeton University Press, Princeton, NJ).
- Lo, I., and S. G. Sapp (2010), "Order aggressiveness and quantity: how are they determined in a limit order market?" *Journal of International Financial Markets, Institutions and Money* **20** (3), 213.
- Luckock, H. (2001), "A statistical model of a limit order market," Sydney University Preprint, available at <http://www.maths.usyd.edu.au/res/AppMaths/Luc/2001-9.pdf>.
- Luckock, H. (2003), "A steady-state model of the continuous double auction," *Quantitative Finance* **3** (5), 385.
- Madhavan, A., D. Porter, and D. Weaver (2005), "Should securities markets be transparent?" *Journal of Financial Markets* **8** (3), 265.
- Maskawa, J. (2007), "Correlation of coming limit price with order book in stock markets," *Physica A* **383** (1), 90.
- Maslov, S. (2000), "Simple model of a limit order-driven market," *Physica A* **278** (3-4), 571.
- Maslov, S., and M. Mills (2001), "Price fluctuations from the order book perspective – Empirical facts and a simple model," *Physica A* **299** (1-2), 234.
- Mike, S., and J. D. Farmer (2008), "An empirical behavioral model of liquidity and volatility," *Journal of Economic Dynamics and Control* **32** (1), 200.
- Mitchell, M. (2009), *Complexity: A guided tour* (Oxford University Press).
- Mittal, H. (2008), "Are you playing in a toxic dark pool?" *The Journal of Trading* **3** (3), 20.
- Mizrach, B. (2008), "The next tick on NASDAQ," *Quantitative Finance* **8** (1), 19.
- Mu, G. H., W. Chen, J. Kertész, and W. X. Zhou (2009), "Preferred numbers and the distributions of trade sizes and trading volumes in the chinese stock market," *The European Physical Journal B* **68** (1), 145.
- Mu, G. H., and W. X. Zhou (2010), "Tests of nonuniversality of the stock return distributions in an emerging market," *Physical Review E* **82** (6), 066103.
- NASDAQ, (2010), Retrieved 20th September, 2011, from [www.nasdaqtrader.com/content/products/services/trading/psx/psxfacts.pdf](http://www.nasdaqtrader.com/content/products/services/trading/psx/psxfacts.pdf).
- Obizhaeva, A., and J. Wang (2005), "Optimal trading strategy and supply/demand dynamics," Working Paper, SSRN eLibrary, available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=752022](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=752022).
- Parlour, C., and D. J. Seppi (2008), "Limit order markets: a survey," in *Handbook of Financial Intermediation and Banking*, edited by A. Thakor and A. Boot (Elsevier).
- Parlour, C. A. (1998), "Price dynamics in limit order markets," *Review of Financial Studies* **11** (4), 789.
- Peng, C. K., S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger (1994), "Mosaic organization of DNA nucleotides," *Physical Review E* **49** (2), 1685.
- Plerou, V., P. Gopikrishnan, X. Gabaix, and H. E. Stanley (2002), "Quantifying stock-price response to demand fluctuations," *Physical Review E* **66** (2), 27104.
- Plerou, V., and H. E. Stanley (2008), "Stock return distributions: Tests of scaling and universality from three distinct stock markets," *Physical Review E* **77** (3), 037101.
- Potters, M., and J. P. Bouchaud (2003), "More statistical properties of order books and price impact," *Physica A* **324**, 133.
- Preis, T., S. Golke, W. Paul, and J. J. Schneider (2006), "Multi-agent-based order book model of financial markets," *Europhysics Letters* **75**, 510.
- Preis, T., S. Golke, W. Paul, and J. J. Schneider (2007), "Statistical analysis of financial returns for a multiagent order book model of asset trading," *Physical Review E* **76** (1), 016108.
- Rinaldo, A. (2004), "Order aggressiveness in limit order book markets," *Journal of Financial Markets* **7** (1), 53.
- Rea, W., L. Oxley, M. Reale, and J. Brown (2009), "Estimators for long-range dependence: an empirical study," arXiv:0901.0762.
- Robinson, P. M., Ed. (2003), *Time series with long memory* (Oxford University Press).
- Roşu, I. (2009), "A dynamic model of the limit order book," *Review of Financial Studies* **22** (11), 4601.
- Roşu, I. (2010), "Liquidity and information in order driven markets," Working paper, SSRN eLibrary, available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1286193](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1286193).
- Sandås, P. (2001), "Adverse selection and competitive market making: empirical evidence from a limit order market," *Review of Financial Studies* **14** (3), 705.
- Seppi, D. J. (1997), "Liquidity provision with limit orders and a strategic specialist," *Review of Financial Studies* **10** (1), 103.
- SETS, (2011), Retrieved 20th September, 2011, from [www.londonstockexchange.com/products-and-services/trading-services/sets/sets.htm](http://www.londonstockexchange.com/products-and-services/trading-services/sets/sets.htm).
- Shephard, N., Ed. (2005), *Stochastic Volatility* (Oxford University Press).
- Slanina, F. (2001), "Mean-field approximation for a limit order driven market model," *Physical Review E* **64** (5), 56136.
- Smith, E., J. D. Farmer, L. Gillemot, and S. Krishnamurthy (2003), "Statistical theory of the continuous double auction," *Quantitative Finance* **3** (6), 481.
- Stanley, H. E., V. Plerou, and X. Gabaix (2008), "A statistical physics view of financial fluctuations: evidence for scaling and universality," *Physica A* **387** (15), 3967.
- Stumpf, M. P. H., and M. A. Porter (2012), "Critical truths about power laws," *Science* **335** (6069), 665.
- Tang, L. H., and G. S. Tian (1999), "Reaction-diffusion-branching models of stock price fluctuations," *Physica A*

- 264** (3–4), 543.
- Taqqu, M. S., V. Teverovsky, and W. Willinger (1995), “Estimators for long-range dependence: an empirical study,” *Fractals* **3** (4), 785.
- Taylor, S. J. (2008), *Modelling financial time series* (World Scientific Publishing).
- Teverovsky, V., M. S. Taqqu, and W. Willinger (1999), “A critical look at lo’s modified r/s statistic,” *Journal of Statistical Planning and Inference* **80** (1–2), 211.
- Thomson-Reuters, (2011), Retrieved 20th September, 2011, from [https://dxtrapub.markets.reuters.com/docs/Matching\\_Rule\\_Book.pdf](https://dxtrapub.markets.reuters.com/docs/Matching_Rule_Book.pdf).
- Toke, I. M. (2011), ““Market making” in an order book model and its impact on the spread,” in *Econophysics of Order-driven Markets: Proceedings of Econophys-Kolkata V*, edited by F. Abergel, B. K. Chakrabarti, C. A., and M. M. (Springer, Milan) pp. 49–64.
- Wyart, M., J. P. Bouchaud, J. Kockelkoren, M. Potters, and M. Vettorazzo (2008), “Relation between bid-ask spread, impact and volatility in order-driven markets,” *Quantitative Finance* **8** (1), 41.
- Xu, L., P. C. Ivanov, K. Hu, Z. Chen, A. Carbone, and H. E. Stanley (2005), “Quantifying signals with power-law correlations: a comparative study of detrended fluctuation analysis and detrended moving average techniques,” *Physical Review E* **71** (5), 051101.
- Zhao, L. (2010), *A model of limit order book dynamics and a consistent estimation procedure*, Ph.D. thesis (Carnegie Mellon University).
- Zhou, B. (1996), “High-frequency data and volatility in foreign-exchange rates,” *Journal of Business & Economic Statistics* **14** (1), 45.
- Zhou, W. X. (2012), “Universal price impact functions of individual trades in an order-driven market,” arXiv:0708.3198.
- Zovko, I., and J. D. Farmer (2002), “The power of patience: a behavioral regularity in limit order placement,” *Quantitative Finance* **2** (5), 387.
- Zumbach, G. (2004), “How trading activity scales with company size in the FTSE 100,” *Quantitative Finance* **4** (4), 441.

## Appendix A: Table of Empirical Studies

Reference	Assets Studied	Date Range	Data Type	Main Points Studied
Aït-Sahalia <i>et al.</i> (2005)	The 30 Dow Jones Industrial Average stocks	19th-23rd and 26th-30th April, 2004	$b(t)$ , $a(t)$ , $n_b(b(t), t)$ and $n_a(a(t), t)$ , and all market orders	Volatility and long range dependence in order flows
Anand <i>et al.</i> (2005)	144 stocks traded on the NYSE	November 1990 to January 1991	All order flows at all prices	Decision between using limit orders or market orders for informed traders
Bandi and Russell (2006)	All stocks in the S&P 100 index	February 2002	$b(t)$ , $a(t)$ , $n_b(b(t), t)$ and $n_a(a(t), t)$ , and all market orders	Volatility
Biais <i>et al.</i> (1995)	The CAC 40, traded on the Paris Bourse	6 trading days in June/July 1991 and 19 trading days in October/November 1991	$N_b(p, t)$ and $N_a(p, t)$ , for $p = 0, 1, 2, 3, 4$ , updated every time one of them changes	Returns, percentage of market orders that match against hidden liquidity, $\bar{N}_b(p)$ , $\bar{N}_a(p)$ , and $s(t)$ (both unconditionally and dependent on time of day), order flow (both unconditionally and dependent on recent order flow and time of day) and state of $L(t)$
Biais <i>et al.</i> (1999)	The CAC 40, traded on the Paris Bourse	19 trading days in October/November 1991, 26 trading days in 1993, and 234 trading days in 1995	Once-per-minute sampling of $b(t)$ and $a(t)$	Whether the evolution of the price process indicated learning on behalf of the market participants during the daily opening auction
Boehmer <i>et al.</i> (2005)	400 stocks traded on the NYSE	January 7th-18th, February 4th-15th, March 4th-15th, April 1st-12th, May 6th-17th, all in 2002	All order flows at all prices in the electronic LOB, plus information about the handling of both electronic and manual (broker-handled) orders	How the introduction of an electronic LOB on the NYSE affected market participants' behaviour
Bortoli <i>et al.</i> (2006)	The 4 most actively traded futures contracts on the Sydney Futures Exchange	September 15th 2000 to June 19th 2001	Every matching, change in $b(t)$ or $a(t)$ , and change in depth available at the best prices (respectively, best three prices) prior to (respectively, after) the change in disseminated market information, timestamped to the nearest second	Whether order flow and the state of the LOB changed when the Sydney Futures Exchange increased the real-time information disseminated to market participants, from only the depths available at $b(t)$ and $a(t)$ to the depths available at the best three prices on each side of the LOB
Bouchaud <i>et al.</i> (2002)	France Telecom, Vivendi, and Total stocks traded on the Paris Bourse	February 2001	All order arrivals at all prices along with their time of arrival, and a list of all orders that were cancelled (but not the time at which they were cancelled)	$\bar{N}_b(p)$ , $\bar{N}_a(p)$ , and the distribution of relative price, order size, $n_b(b(t), t)$ , and $n_a(a(t), t)$
Bouchaud and Potters (2003)	France Telecom Stock, traded on the Paris Bourse (but similar results reported for other unnamed liquid French and British stocks)	Trading days during 2001 and 2002	$b(t)$ , $a(t)$ , $n_b(b(t), t)$ and $n_a(a(t), t)$ , recorded once every time any of these changes and timestamped to the nearest second, and all market orders, timestamped to the nearest second	How order flow affected prices

Cao <i>et al.</i> (2008)	100 largest stocks traded on the Australian Stock Exchange	March 2000	All order arrivals and cancellations at all prices, timestamped to the nearest 0.01 seconds	How the state of the LOB affected order flow
Chakraborti <i>et al.</i> (2011b)	Four stocks traded on the Paris Bourse	All trading days between 1st October 2007 and 30th May 2008	All market orders, plus five highest priority active orders on each side of the LOB	Whether the traditional stylized facts were present in the data
Challet and Stinchcombe (2001)	Four stocks traded on the Island ECN (on NASDAQ)	Not specified	15 highest priority active orders on each side of the LOB, updated every time the list changed	Order flow rates, autocorrelation of order flow rates, diffusion of active orders (i.e., cancellation of an active order immediately followed by resubmission at a neighbouring price), instantaneous price impact; distribution of order size, lifetime of limit orders, relative price for incoming orders
Cont <i>et al.</i> (2010)	Sky Perfect Communications stock, traded on the Tokyo Stock Exchange	Not specified	$N_b(p, t)$ and $N_a(p, t)$ for the five smallest relative prices with nonzero depth available, updated every whenever either one changed, and all market orders	Arrival rates of market orders and arrival and cancellation rates of limit orders
Cont <i>et al.</i> (2011)	50 stocks chosen at random from the S&P 500, traded on the NYSE	All 21 trading days in April 2010	$n_b(b(t), t)$ and $n_a(a(t), t)$ , updated whenever either one changed, with a timestamp rounded to the nearest second, and all market orders	Relationship between order flow imbalance and price impact
Dufour and Engle (2000)	18 of the most frequently traded stocks on the NYSE	62 trading days between 1st November 1990 and 31st January 1991	$b(t)$ and $a(t)$ , updated every time they change, and all market orders	Relationship between market order inter-arrival times and price impact
Eisler <i>et al.</i> (2010)	14 randomly selected stocks traded on NASDAQ	The 53 trading days between 3rd March 2008 and 19th May 2008	$b(t)$ , $a(t)$ , $n_b(b(t), t)$ and $n_a(a(t), t)$ , updated every time any of them change	The price impact of market order submissions and limit order submissions and cancellations
Ellul <i>et al.</i> (2003)	The 50 most actively traded stocks and 98 randomly chosen stocks on the NYSE	The 5 trading days between 30th April 2001 and 5th May 2001	All market order submissions and all limit order submissions and cancellations, timestamped to the nearest second	What factors market participants used when choosing the price of their orders
Engle and Patton (2004)	100 randomly selected stocks traded on the NYSE	18 months of data, no date range specified	$b(t)$ and $a(t)$ , updated every time they change, and all market orders	$s(t)$ and how price impact varied according to how frequently trades occur for a specific stock
Farmer and Lillo (2004)	3 stocks traded on the LSE and 3 stocks traded on the NYSE	May 2000 to December 2002 for the LSE stocks and 1995-1996 for the NYSE stocks	All order flows for the LSE; $b(t)$ and $a(t)$ , updated every time they change, and all market orders, for the NYSE	Price impact of individual market orders, and distribution of order sizes for market orders

Farmer <i>et al.</i> (2005)	11 stocks traded on the LSE	All 434 trading days between 1st August 1998 and 30th April 2000	All market order submissions and all limit order submissions and cancellations	Goodness of fit of the predictions regarding mean spread and price diffusion of the Smith <i>et al.</i> (2003) model to data, and mean instantaneous mid-price logarithmic return impact as a function of market order size
Field and Large (2008)	Short Sterling, Euribor, Eurodollar, and 2-Year US Treasury Note futures	23rd November to 11th December 2006 and 16th to 20th April 2007	$b(t)$ , $a(t)$ , $n_b(b(t), t)$ and $n_a(a(t), t)$ , updated every time any of them change	Order flow rates and $n_b(b(t), t)$ and $n_a(a(t), t)$ in markets where $s(t) = \delta p$
Gode and Sunder (1993)	Laboratory experiment with human beings and computerized zero-intelligence traders	N/A	All order flows at all prices	Relative applicability of perfect-rationality and zero-intelligence assumptions, and emergence of seemingly rational behaviour when aggregating across irrational individuals
Gopikrishnan <i>et al.</i> (2000)	1000 largest stocks traded in the US	1994-1995	$a(t)$ , $b(t)$ , and all market orders	Price impact as a function of trade imbalance count and trade imbalance size, and distribution and autocorrelation of trade imbalance count and trade imbalance size
Gu <i>et al.</i> (2008a)	Aggregation of 23 stocks traded on the Shenzhen Stock Exchange	The whole of 2003	All order flows at all prices	Distribution of mid-price returns on various $\tau$ second timescales and various event-by-event timescales
Gu <i>et al.</i> (2008b)	Aggregation of 23 stocks traded on the Shenzhen Stock Exchange	The whole of 2003	All order flows at all prices	Distribution of relative prices of incoming orders, and whether this is conditional on $s(t)$ or volatility
Gu <i>et al.</i> (2008c)	23 stocks traded on the Shenzhen Stock Exchange	The whole of 2003	All order flows at all prices	$N_b(p)$ , $N_a(p)$ , and changes in relative depth profiles through time
Gu and Zhou (2009a)	23 stocks traded on the Shenzhen Stock Exchange	The whole of 2003	All order flows at all prices	Autocorrelation of relative prices of incoming orders
Hall and Hautsch (2006)	The 5 most liquid stocks traded on the Australian Stock Exchange	July to August 2002	All order flows at all prices	Whether the distribution of relative prices of incoming orders was conditional on the state of the LOB, volatility, and recent order flows
Harris and Hasbrouck (1996)	144 randomly selected stocks traded on the NYSE	November 1990 to January 1991	All order flows at all prices	Analysis of performance measures aiding decision-making between limit orders versus market orders
Hasbrouck and Saar (2002)	The 300 largest equities on NASDAQ, traded on Island ECN	1st October to 31st December 1999	All order flows at all prices	How volatility was related to order flow and the state of the LOB, and how order fill probabilities and mean time to execution varied with volatility

Hautsch and Huang (2009)	The 30 most frequently traded stocks on Euronext Amsterdam	All trading days between 1st August and 30th September, 2008	$N_b(p, t)$ and $N_a(p, t)$ for $p = 0, 1, 2$ , updated every whenever either one changed, and a record of all trades that actually occurred, timestamped to the nearest millisecond	Market impact of incoming limit orders
Hendershott and Jones (2005)	3 exchange traded funds on Island ECN	16th August to 31st October 2002	For activity on Island: for the first part of the data, $b(t)$ , $a(t)$ , $n_b(b(t), t)$ and $n_a(a(t), t)$ , updated every time any of them change, and all market orders; for the second part of the data, only market orders; for activity not on Island, $b(t)$ , $a(t)$ , $n_b(b(t), t)$ and $n_a(a(t), t)$ , updated every time any of them change, and all market orders, for the entire data period	The effect that showing $L(t)$ to market participants had on price series
Hendershott <i>et al.</i> (2011)	943 stocks traded on the NYSE	February 2001 to December 2005	$b(t)$ , $a(t)$ , $n_b(b(t), t)$ and $n_a(a(t), t)$ , updated every time any of them change	The effects of algorithmic trading on $L(t)$
Hollifield <i>et al.</i> (2004)	The Ericsson stock, traded on the Stockholm Stock Exchange	The 59 trading days between 3rd December 1991 and 2nd March 1992	All order flows at all prices	Whether market participants' actions could be explained by a cut-off strategy based on their private valuation of the asset
Hollifield <i>et al.</i> (2006)	3 stocks traded on the Vancouver Stock Exchange	May 1990 to November 1993	All order flows at all prices	Distribution of traders' personal valuations, inferred from their actions
Kempf and Korn (1999)	DAX futures contracts, traded on the German Futures and Options Exchange	17th September 1993 to 15th September 1994	$b(t)$ , $a(t)$ and all market orders	Permanent price impact, as a function of several measures of trade imbalance, over 1 minute time horizons
Lillo and Farmer (2004)	20 stocks traded on the LSE	1999 to 2002	All order flows at all prices	Autocorrelation of order sizes, mid prices, $n_b(b(t), t)$ , $n_a(a(t), t)$ , and order type (buy or sell) for arriving LOs, arriving MOs, and cancelled LOs
Lillo <i>et al.</i> (2005)	20 stocks traded on the LSE	May 2000 to December 2002	All LOB order flows and all off-book trades for the same stocks	Effects of order splitting and hidden liquidity on observed order flows
Lillo (2007)	Astrazeneca Stock, traded on the LSE	May 2000 to December 2002	Order arrivals, partitioned by who submitted them	Distribution of relative prices for incoming limit orders from specified market participants
Lo and Sapp (2010)	Deutsche Mark/US dollar and Canadian dollar/US dollar currency pairs	5th October to 10th October 1997 for Deutsche Mark/US dollar; 1st May to 30th June 2005 for Canadian dollar/US dollar	All order flows at all prices	How market participants chose the size and relative prices of their orders

Madhavan <i>et al.</i> (2005)	109 stocks traded via a LOB and 240 stocks traded by floor traders on the Toronto Stock Exchange	March and May, 1990	For March, $b(t)$ , $a(t)$ , $n_a(a(t), t)$ , and $n_b(b(t), t)$ ; for May, $b(t)$ , $a(t)$ , $N_b(p, t)$ , and $N_a(p, t)$ , for $p = 0, 1, 2, 3, 4$ , and all market orders; all floor-trader trades for both months	How real-time disclosure of more information about the depth profile affected market participants' behaviour
Maskawa (2007)	13 stocks traded on the LSE	July to December 2004	All order flows at all prices	Distribution of relative prices for incoming limit orders, and whether this distribution was affected by the state of $L(t)$
Maslov and Mills (2001)	Cisco Systems, Broadcom Corporation, and JDS Uniphase Corporation stocks traded on NASDAQ	30th June 2000 for Cisco Systems stock; 3rd July for Broadcom Corporation stock; and 5th, 6th, and 11th July for JDS Uniphase Corporation stock	$b(t)$ , $a(t)$ , $N_b(p, t)$ and $N_a(p, t)$ for $p = 0, 1, 2, 3$ , and all market orders	Distribution of order sizes, $n_b(b(t), t)$ , $(n_a(a(t), t))$ , depth profiles, instantaneous price impact
Mike and Farmer (2008)	25 stocks traded on the LSE	May 2000 to December 2002	All order flows at all prices	Relative prices of incoming orders, autocorrelation of order type in order flows, order cancellations
Mizrach (2008)	The 4 largest stocks on NASDAQ; 95 of the "NASDAQ 100" stocks; and 87 other randomly chosen smaller NASDAQ stocks	December 2002	All order flows at all prices	How $L(t)$ affected the next change in $b(t)$ or $a(t)$
Mu <i>et al.</i> (2009)	22 stocks traded on the Shenzhen Stock Exchange	The whole of 2003	All order flows at all prices	Distribution of market order sizes
Mu and Zhou (2010)	978 stocks traded on the Shenzhen Stock Exchange	January 2004 to June 2006	$b(t)$ and $a(t)$ , updated once every 6 to 8 seconds	Distribution of mid-price logarithmic returns for stocks in emerging markets, and how this varied with time window and market capitalization of the stock studied
Plerou <i>et al.</i> (2002)	The 116 most frequently traded US stocks	1994 to 1995	$b(t)$ , $a(t)$ , $n_b(b(t), t)$ and $n_a(a(t), t)$ , and all market orders	Price impact as a function of trade imbalance count and trade imbalance size, over a variety of time horizons
Plerou and Stanley (2008)	1000 major US stocks; 85 of the FTSE 100 stocks (traded on the LSE); 13 of the CAC 40 stocks (traded on the Paris Bourse); 422 stocks from the Center for Research in Security Prices (CRSP)	1994-1995 for US stocks; 2001-2002 for LSE stocks; 3rd Jan 1995-22nd Oct 1999 for Paris Bourse stocks; Jan 1962-Dec 1996 for CRSP database stocks	All market orders	Distribution of mid-price returns and number of arriving market orders, and whether they varied according to market capitalization or industry sector, on various $\tau$ second timescales
Potters and Bouchaud (2003)	Exchange traded funds that track NASDAQ and the S&P 500, and the Microsoft stock	1st June to 15th July, 2002	All order flows at all prices	Distribution of relative prices, relative depth profiles, arrival and cancellation rates, instantaneous price impact

Rinaldo (2004)	15 stocks traded on the Swiss Stock Exchange	March and April 1997	$b(t)$ , $a(t)$ , $n_b(b(t), t)$ and $n_a(a(t), t)$ , and all market orders	How volatility, recent order flow, and the state of $L(t)$ affected order flow, intraday patterns in spread and volatility, symmetry between the buy and sell sides of the LOB
Sandås (2001)	10 stocks traded on the Stockholm Stock Exchange	59 trading days between 3rd December 1991 and 2nd March 1992	All order flows at all prices	Whether the depth profile supported hypotheses about how market participants make decisions related to order submissions and cancellations
Toke (2011)	3 stocks from the CAC40, 3 month Euribor futures, and FTSE 100 futures	10th September 2009 to 30th September 2009	$N_b(p, t)$ , and $N_a(p, t)$ , for $p = 0, 1, 2, 3, 4$ , updated whenever either one changed, timestamped to the millisecond	Whether Hawkes processes provided a better explanation of order flows than do Poisson processes
Wyart <i>et al.</i> (2008)	The 68 most liquid stocks on the Paris Bourse, small tick index futures contracts, and the 155 most actively traded stocks on the NYSE	2002 for the Paris Bourse, 2005 for the small tick futures and NYSE stocks	$b(t)$ , $a(t)$ , $n_b(b(t), t)$ and $n_a(a(t), t)$ , and all market orders	How the profit of a market maker trading in a LOB depended on $s(t)$ , and price impact
Zhao (2010)	Crude oil futures contracts, traded on the International Petroleum Exchange	17th October 2005	$N_b(p, t)$ , and $N_a(p, t)$ , for $p = 0, 1, 2, 3, 4$ , updated whenever either one changed, and all market orders, timestamped to the nearest second	Order flow rates
Zhou (1996)	Deutsche Mark/US dollar, US dollar/Yen and Deutsche Mark/Yen currency pairs, traded on Reuters	1st October 1992 to 30th September 1993	$b(t)$	Volatility
Zhou (2012)	23 stocks traded on the Shenzhen Stock Exchange (although 1 is later removed as its price was reported to be manipulated in the data)	The whole of 2003	All order flows at all prices	Instantaneous price impact of individual orders
Zovko and Farmer (2002)	50 stocks traded on the LSE	1st August 1998 to 31st April 2000	Relative prices of incoming limit orders	Relative prices of incoming orders, autocorrelation of order type in order flows, and volatility